CHAPTER II

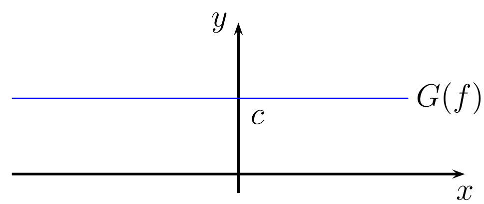
CONTINUOUS FUNCTIONS OF A REAL VARIABLE

§5. CONTINUITY

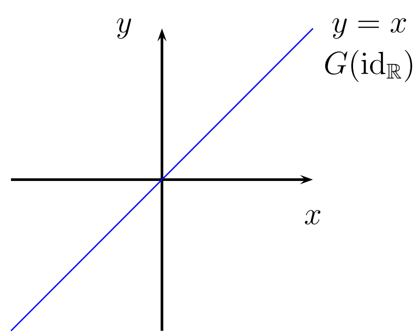
In this section we study real (valued) functions[1] on subsets of $\mathbb{R}$, i.e., maps $f\colon D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$. Recall that the *graph* ⟨*Graph*⟩ of f is defined as the following subset of $\mathbb{R}^2$:

$$G(f) := \{(x, f(x)) \in \mathbb{R}^2 : x \in D\}.$$

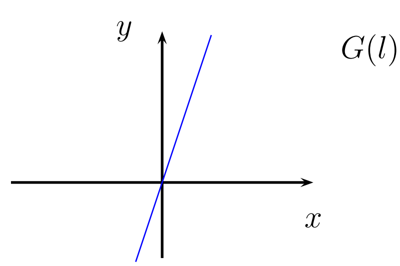
5.1. Examples: 1) Let $c \in \mathbb{R}$ arbitrary, then $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) := c$ for all $x \in \mathbb{R}$ defines a *constant function*.



2) The *identity map* ⟨*identische Abbildung*⟩ on $\mathbb{R}$ is given by $\mathrm{id}_{\mathbb{R}}\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x$.

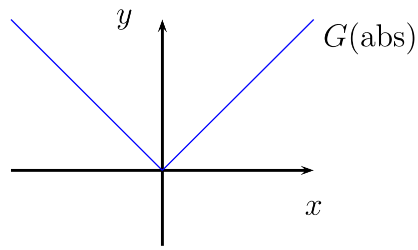


Slightly more general are linear functions $l\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto a \cdot x$, where $a \in \mathbb{R}$ gives the slope of the graph:

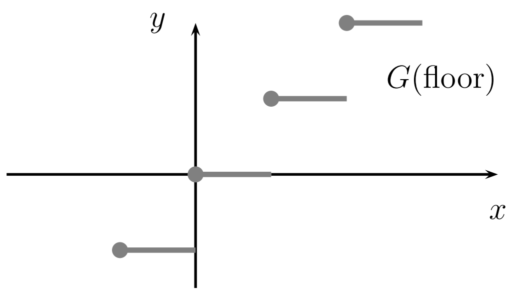


[1]The mathematical term, “function” (from the Latin functio, meaning performance, execution) was first used by Leibniz in 1694 to describe curves.

3) The *absolute value (function)* ⟨*Betragsfunktion oder Absolutbetrag*⟩ is defined by $\mathrm{abs}\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$.

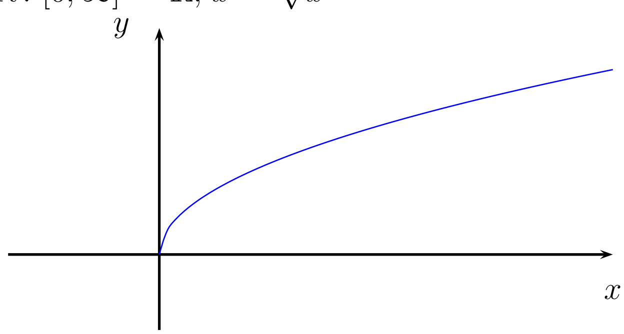


4) $\mathrm{floor}\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \lfloor x \rfloor$, where (as on page 44) $\lfloor x \rfloor = \max\{n \in \mathbb{Z} \colon n \leq x\}$.



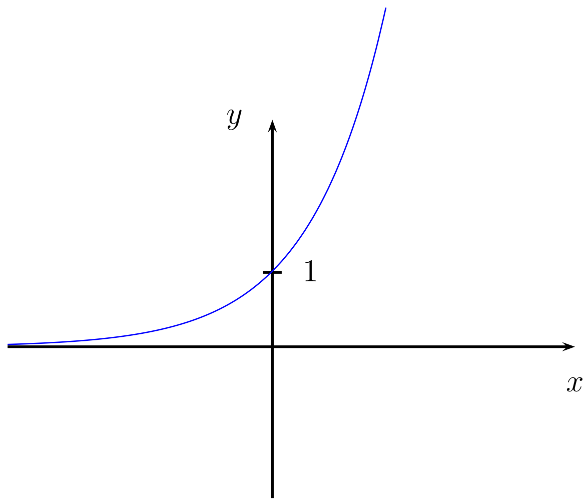
The floor function is sometimes called *Gauß bracket*[2] ⟨*Gaußklammer*⟩ and the values are also denoted by $[x]$ $(x \in \mathbb{R})$.

5) The square root $\mathrm{sqrt}\colon [0, \infty] \to \mathbb{R}$, $x \mapsto \sqrt{x}$



[2] Carl Friedrich Gauß (1777–1855) [kaʁl ˈfriːtrɪç gaʊs], one of the most outstanding German mathematicians

6) The exponential function $\exp\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \exp(x)$ as defined in 4.23.

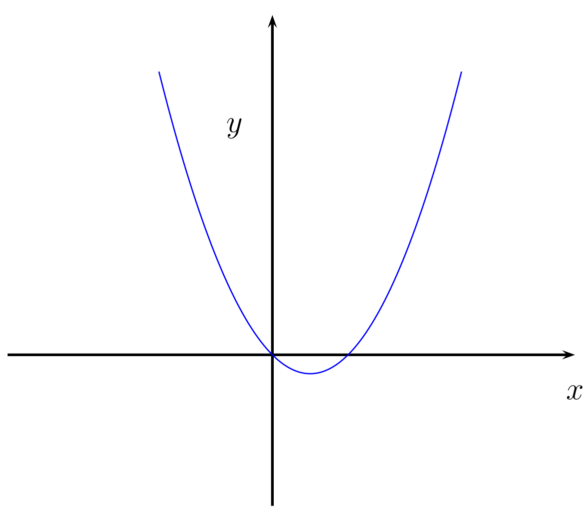


7) *Polynomial functions* ⟨Polynomfunktionen⟩: Let $m \in \mathbb{N}$ and $a_0, a_1, \dots, a_m \in \mathbb{R}$. We define

$$p\colon \mathbb{R} \to \mathbb{R} \quad \text{by} \quad p(x) := a_m x^m + a_{m-1}x^{m-1} + \dots a_1 x + a_0 \qquad \forall x \in \mathbb{R}.$$

The constants $a_0, \dots a_m$ are called the *coefficients* ⟨Koeffizienten⟩ of the polynomial function. If $a_m \neq 0$ then p is said to be of *degree* m ⟨vom Grad m⟩.

For example, when $m = 2$ and $a_0 = 0$, $a_1 = -1$, $a_2 = 1$ we obtain $p(x) = x^2 - x$



8) *Rational functions* ⟨rationale Funktionen⟩: Let p and q be polynomial functions, that is

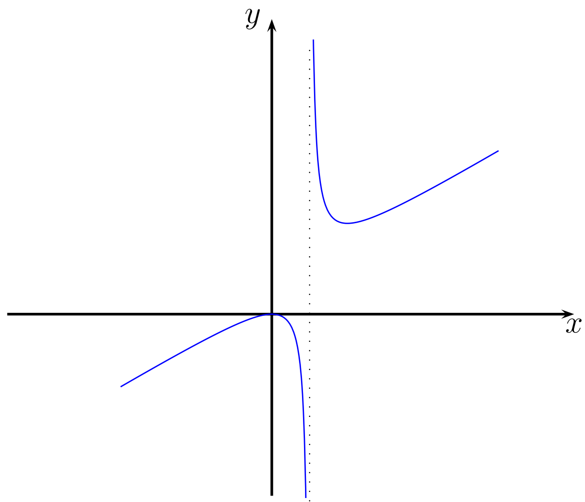
$$p(x) = a_m x^m + \dots + a_1 x + a_0 \quad \text{and} \quad q(x) = b_n x^n + \dots + b_1 x + b_0$$

with given coefficients $a_0, \dots, a_m, b_0, \dots b_n \in \mathbb{R}$. Then a rational function is the quotient function with domain $D := \{x \in \mathbb{R} : q(x) \neq 0\}$, defined by

$$r\colon D \to \mathbb{R}, \quad x \mapsto \frac{p(x)}{q(x)}.$$

Note that polynomial functions are just rational functions with denominator $q \equiv 1$.

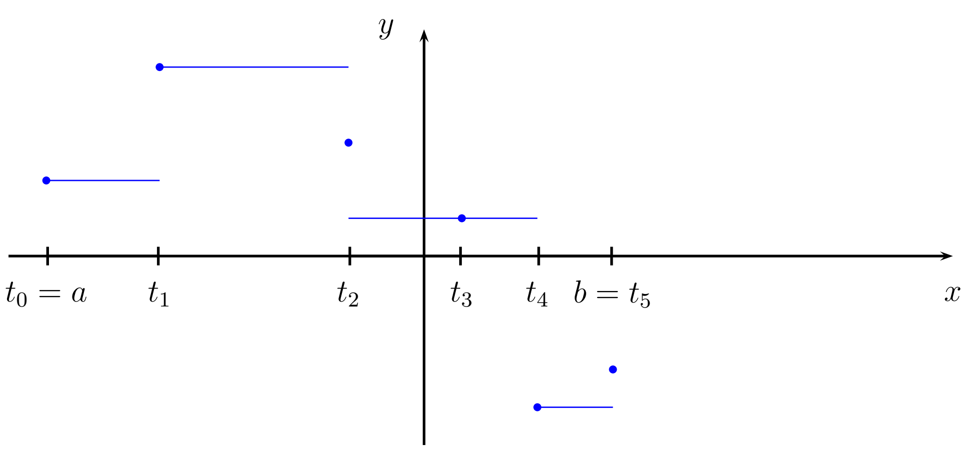
For example, here is the graph of the rational function $r\colon \mathbb{R} \setminus \{1\} \to \mathbb{R}$, $r(x) = x^2/(x-1)$



9) *Simple functions (or step functions)* ⟨Treppenfunktionen⟩: Let $a, b \in \mathbb{R}$ with $a < b$. A function $\varphi\colon [a, b] \to \mathbb{R}$ is called a simple function (or step function), if there is a finite partition $a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$ of the interval $[a, b]$ and coefficients $c_1, \ldots c_n \in \mathbb{R}$ such that

$$\varphi(x) = c_k \qquad \text{when } x \in]t_{k-1}, t_k[\quad (1 \leq k \leq n).$$

Therefore φ is constant on each open subinterval $]t_{k-1}, t_k[$ $(1 \leq k \leq n)$ but the finitely many values $\varphi(t_k)$ $(0 \leq k \leq n)$ are arbitrary.



Note that the restriction floor $|_{[a,b]}$ of the floor function provides an example of a simple function.

10) The *characteristic function of* $\mathbb{Q}$ ⟨*charakteristische Funktion von* $\mathbb{Q}$⟩ or *Dirichlet function*[3] is given by

$$\mathbb{1}_{\mathbb{Q}}\colon \mathbb{R} \to \mathbb{R}, \quad \mathbb{1}_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q}. \end{cases}$$

In this case the graph is

$$G(\mathbb{1}_{\mathbb{Q}}) = \{(q,1) : q \in \mathbb{Q}\} \cup \{(s,0) : s \in \mathbb{R} \setminus \mathbb{Q}\},$$

which would be somewhat hard to depict ...

5.2. Review of basic operations with functions:

Let $f, g\colon D \to \mathbb{R}$ be functions on $D \subseteq \mathbb{R}$ and $\lambda \in \mathbb{R}$.

- Then the functions

  $$f + g\colon D \to \mathbb{R}, \qquad \lambda f\colon D \to \mathbb{R}, \qquad f \cdot g\colon D \to \mathbb{R}$$

  are defined in terms of the corresponding pointwise operations (with real numbers) for all $x \in D$ by

  $$\begin{aligned} (f+g)(x) &:= f(x) + g(x), \\ (\lambda f)(x) &:= \lambda \cdot f(x), \\ (f \cdot g)(x) &:= f(x) \cdot g(x). \end{aligned}$$

  Remark: It is easy to check that the set $\mathcal{F}(D) := \{f\colon D \to \mathbb{R}\}$ of all real valued functions on the set D together with the addition and scalar multiplication as defined by the first two lines above forms a vector space over $\mathbb{R}$.

- Let $D' := \{x \in D : g(x) \neq 0\}$. The quotient function is defined by

  $$\frac{f}{g}\colon D' \to \mathbb{R}, \quad x \to \frac{f(x)}{g(x)}.$$

- Let $E \subseteq \mathbb{R}$ such that $f(D) \subseteq E$ and $h\colon E \to \mathbb{R}$. Recall that the composition of f and h is given by

  $$h \circ f\colon D \to \mathbb{R}, \quad (h \circ f)(x) := h(f(x)) \qquad \forall x \in D.$$

[3] Johann Peter Gustav Lejeune Dirichlet (1805–1859) [ˈjoːhan ˈpeːtɐ ˈɡʊstaf ləˈʒœn diʀiˈkle], German mathematician with Belgish origins (the French words Lejeune Dirichlet literally mean “the young chap from Richelet”)

Examples: 1) If $q\colon \mathbb{R} \to \mathbb{R}$, $q(x) = x^2$, then $q = \mathrm{id} \cdot \mathrm{id}$.

2) More generally, if p is a polynomial function, given by

$$p(x) = a_m x^m + \ldots a_1 x + a_0,$$

then

$$p = a_m \cdot \underbrace{(\mathrm{id} \cdot \mathrm{id} \cdots \mathrm{id})}_{m \text{ factors}} + \ldots a_1 \cdot \mathrm{id} + \mathbb{1},$$

where $\mathbb{1}$ denotes the constant function $\mathbb{1}(x) = 1$ for all $x \in \mathbb{R}$.

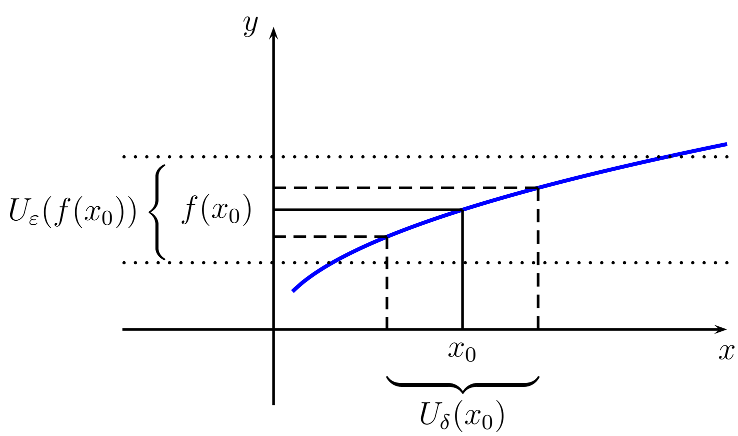
3) With q as in example 1) we have $\mathrm{abs} = \mathrm{sqrt} \circ q$, since for all $x \in \mathbb{R}$

$$\big(\mathrm{sqrt} \circ q\big)(x) = \sqrt{x^2} = |x| = \mathrm{abs}(x).$$

5.3. Continuity ⟨*Stetigkeit*⟩:

The notion of continuity of a function is a precise way to express an intuitive requirement, which is often implicitly made in model applications: Namely, that small perturbations of a function argument should not result in extreme changes of the function values.

How to specify such a property for a given function f near a point x_0 of its domain? It might seem practically desirable to first prescribe the acceptable tolerance around the value $f(x_0)$ and then to look for a safety interval around the argument x_0 on which function values near $f(x_0)$ within tolerance are guaranteed. If the tolerance is given in terms of an interval $]f(x_0) - \varepsilon, f(x_0) + \varepsilon[$ with $\varepsilon > 0$ and the safety interval is sought in the form $]x_0 - \delta, x_0 + \delta[$ with $\delta > 0$ we obtain the following picture:



By requiring that for every tolerance $\varepsilon > 0$ — chosen arbitrarily small — an appropriate safety guard $\delta > 0$ can (in principle) be found we arrive at the notion of continuity.

DEFINITION: Let $x_0 \in D \subseteq \mathbb{R}$ and $f\colon D \to \mathbb{R}$. The function f is *continuous* ⟨*stetig*⟩ at x_0 if

$$(5.1) \qquad \forall \varepsilon > 0\, \exists \delta > 0: \quad \forall x \in D: \quad |x - x_0| < \delta \quad \Longrightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

Equivalently, upon recalling that $]x_0 - \delta, x_0 + \delta[\, = U_\delta(x_0)$ and $]f(x_0) - \varepsilon, f(x_0) + \varepsilon[\, = U_\varepsilon(f(x_0))$, we can define the continuity of f at x_0 in terms of neighborhoods:

$$\forall \varepsilon > 0\, \exists \delta > 0: \quad f\big(U_\delta(x_0) \cap D\big) \subseteq U_\varepsilon\big(f(x_0)\big).$$

The function f is said to be *continuous (on D)* if it is continuous at each point in D. If f is not continuous at a point $b \in D$ then f is said to be *discontinuous* ⟨*unstetig*⟩ at b.

EXAMPLES: 1) Clearly, a constant function f is continuous (at every point x_0 in its domain), since $f(x) - f(x_0) = 0$ and therefore (5.1) is satisfied for all $\varepsilon > 0$ and $\delta > 0$ arbitrary.

2) Every linear function $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto ax$, is continuous (at every $x_0 \in \mathbb{R}$): If $a = 0$ this is clear from Example 1), hence consider $a \neq 0$. Let $\varepsilon > 0$. From the preparatory observation $|f(x) - f(x_0)| = |a||x - x_0|$ we learn that we can simply choose $\delta := \varepsilon/|a|$ to achieve (5.1): Indeed, if $|x - x_0| < \delta = \varepsilon/|a|$ then

$$|f(x) - f(x_0)| = |a|\,|x - x_0| < |a|\,\delta = \varepsilon.$$

3) The exponential function $\exp\colon \mathbb{R} \to \mathbb{R}$ is continuous: Let $x_0 \in \mathbb{R}$ and $\varepsilon > 0$. By the properties of the exponential function we have

$$|\exp(x) - \exp(x_0)| = \exp(x_0)\,|\exp(x - x_0) - 1|,$$

where $\exp(x_0) > 0$. From (4.5) we obtain for $|x - x_0| \leq 1$ that

$$|\exp(x - x_0) - 1| \leq 2|x - x_0|.$$

Thus, putting $\delta := \min(1, \frac{\varepsilon}{2\exp(x_0)})$ and combining the above inequalities we obtain for all x with $|x - x_0| < \delta$ the required estimate

$$|\exp(x) - \exp(x_0)| \leq 2\exp(x_0)\,|x - x_0| < 2\exp(x_0)\,\delta = \varepsilon.$$

4) $\mathrm{abs}\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$, is continuous: Let $x_0 \in \mathbb{R}$ and $\varepsilon > 0$. Put $\delta := \varepsilon$ then we have for all $x \in U_\delta(x_0)$

$$|\mathrm{abs}(x) - \mathrm{abs}(x_0)| = ||x| - |x_0|| \leq |x - x_0| < \delta = \varepsilon.$$

5) The Dirichlet function $\mathbb{1}_\mathbb{Q}$ [Example 5.1, 10)] is discontinuous at every point in $\mathbb{R}$: Let $x_0 \in \mathbb{R}$ and put $\varepsilon = 1/2$.

If $x_0 \in \mathbb{R} \setminus \mathbb{Q}$ then $\mathbb{1}_{\mathbb{Q}}(x_0) = 0$. By the density of $\mathbb{Q}$ in $\mathbb{R}$, for every $\delta > 0$ we might choose the interval $U_\delta(x_0) =]x_0 - \delta, x_0 + \delta[$ will always contain some (in fact, many) rational number(s) r [cf. 0.7 or the Corollary in **??**]. In other words, we can find r with $|r - x_0| < \delta$ but

$$|\mathbb{1}_{\mathbb{Q}}(r) - \mathbb{1}_{\mathbb{Q}}(x_0)| = |1 - 0| = 1 \geq \frac{1}{2} = \varepsilon.$$

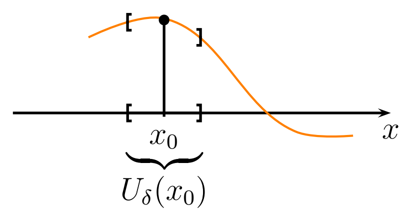
If $x_0 \in \mathbb{Q}$ then $\mathbb{1}_{\mathbb{Q}}(x_0) = 1$. Recall that also $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$ [cf. 0.7]. Hence for every $\delta > 0$ we can find $s \in U_\delta(x_0) \cap (\mathbb{R} \setminus \mathbb{Q})$, which implies

$$|\mathbb{1}_{\mathbb{Q}}(s) - \mathbb{1}_{\mathbb{Q}}(x_0)| = |0 - 1| = 1 \geq \frac{1}{2} = \varepsilon$$

while $|s - x_0| < \delta$.

Knowing that a specific value of a continuous function has positive distance to a certain real number c already guarantees that the function values will stay away from c in a whole neighborhood. In the following statement we formulate this for the special case with $c = 0$. This can easily be adapted to the case $c \neq 0$ by a simple translation of the function graph.

5.4. Lemma: Let $f \colon D \to \mathbb{R}$ be continuous at $x_0 \in D$ and assume that $f(x_0) \neq 0$. Then there is $\delta > 0$ such that for all $x \in U_\delta(x_0) \cap D$ we have $f(x) \neq 0$.



Proof. Put $\varepsilon := |f(x_0)|/2$. Then clearly $\varepsilon > 0$ and by continuity there exists some $\delta > 0$ such that for all $x \in D$ with $|x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \varepsilon = |f(x_0)|/2$. Therefore $x \in U_\delta(x_0) \cap D$ implies

$$|f(x)| = |f(x_0) + f(x) - f(x_0)| \geq |f(x_0)| - |f(x) - f(x_0)| > |f(x_0)| - \varepsilon = \frac{|f(x_0)|}{2} > 0.$$

□

5.5. Continuity test by sequences:

THEOREM: Let $a \in D \subseteq \mathbb{R}$ and $f \colon D \to \mathbb{R}$. The following are equivalent:

(i) f is continuous at a.

(ii) For every sequence (x_n) with $x_n \in D$ we have: if $\lim x_n = a$ then $\lim f(x_n) = f(a)$.

Proof. (i) $\Rightarrow$ (ii): Let $x_n \in D$ $(n \in \mathbb{N})$ with $\lim x_n = a$ and let $\varepsilon > 0$. Choose $\delta > 0$ such that the continuity condition (5.1) is satisfied. There exists $n_0 \in \mathbb{N}$ such that $|x_n - a| < \delta$ holds for all $n \geq n_0$. Thus (5.1) implies

$$|f(x_n) - f(a)| < \varepsilon \quad \forall n \geq n_0,$$

which proves that $\lim f(x_n) = f(a)$.

(ii) $\Rightarrow$ (i): (proof by contradiction) Assume that (ii) holds but (5.1) is false. That is,

$$\exists \varepsilon > 0 \; \forall \delta > 0 : \exists x \in U_\delta(a) \cap D : f(x) \notin U_\varepsilon(f(a)).$$

In particular, with this same $\varepsilon > 0$, we can choose the δ-values to be $1/n$ ($n \in \mathbb{N}$, $n \geq 1$) successively and obtain:

$$\forall n \in \mathbb{N}, n \geq 1 : \exists x_n \in D : |x_n - a| < \frac{1}{n}, \text{ but } |f(x_n) - f(a)| \geq \varepsilon.$$

Therefore $\lim x_n = a$ whereas $f(x_n) \not\to f(a)$ $(n \to \infty)$ — a contradiction ↯ . □

EXAMPLE: The function floor: $\mathbb{R} \to \mathbb{R}$, $x \mapsto \lfloor x \rfloor$ is continuous in $\mathbb{R} \backslash \mathbb{Z}$ and discontinuous in all points $a \in \mathbb{Z}$.

If $a \in \mathbb{Z}$ then $\lfloor a \rfloor = a$ and the sequence $x_n := a - \frac{1}{n}$ $(n \geq 1)$ has $\lim x_n = a$ but $\lim \lfloor x_n \rfloor = \lim(a-1) = a - 1 \neq \lfloor a \rfloor$.

If $a \in \mathbb{R} \setminus \mathbb{Z}$ then $\lfloor a \rfloor < a < \lfloor a \rfloor + 1$. Hence for every sequence (x_n) with $\lim x_n = a$ there exists some n_0 such that $\lfloor a \rfloor < x_n < \lfloor a \rfloor + 1$ when $n \geq n_0$. Therefore $\lfloor x_n \rfloor = \lfloor a \rfloor$ for all $n \geq n_0$, in particular $\lim \lfloor x_n \rfloor = \lfloor a \rfloor$.

5.6. Basic operations and continuity: The following results show that we do not leave the class of continuous functions when applying the basic operations summarized in 5.2 to continuous functions. In other words, we can generate many “new” continuous functions from a set of given continuous functions simply by pointwise addition, scalar multiplication, multiplication, division (when the denominator does not vanish), and composition (where the images and domains match appropriately).

PROPOSITION: (i) Let $a \in D \subseteq \mathbb{R}$ and $\lambda \in \mathbb{R}$. If $f, g \colon D \to \mathbb{R}$ are continuous at a then also

$$f + g \colon D \to \mathbb{R}, \qquad \lambda f \colon D \to \mathbb{R}, \qquad f \cdot g \colon D \to \mathbb{R}$$

are continuous at a. Furthermore, if $a \in D' := \{x \in D : g(x) \neq 0\}$ then

$$\frac{f}{g} \colon D' \to \mathbb{R}$$

is continuous at a.

(ii) Let $D \subseteq \mathbb{R}$, $E \subseteq \mathbb{R}$ and $f \colon D \to \mathbb{R}$, $g \colon E \to \mathbb{R}$ such that $f(D) \subseteq E$. If f is continuous at $a \in D$ and g is continuous at $b := f(a) \in E$ then the composition $g \circ f \colon D \to \mathbb{R}$ is continuous at a.

Proof. (i) Let (x_n) be a sequence in D, respectively D', such that $x_n \to a$. Then by the corresponding properties of basic operations with convergent sequences in 2.10 we obtain that

$$(f+g)(x_n) = f(x_n) + g(x_n) \to f(a) + g(a) = (f+g)(a) \quad (n \to \infty)$$

and similarly for the other types of operations. Thus Theorem 5.5 proves continuity at a.

(ii) Let (x_n) be a sequence in D such that $x_n \to a$. Since f is continuous at a we have $y_n := f(x_n) \to f(a) = b$. Continuity of g at b implies $g(y_n) \to g(b)$. Therefore

$$\lim_{n\to\infty} (g \circ f)(x_n) = \lim g(f(x_n)) = \lim g(y_n) = g(b) = g(f(a)) = (g \circ f)(a)$$

and again by Theorem 5.5 the continuity at a follows. □

COROLLARY: Polynomial functions and rational functions are continuous (on their respective domains).

Proof. By 5.3, Examples 1) and 2), constant functions and the identity map $\mathrm{id}\colon \mathbb{R} \to \mathbb{R}$ are continuous. In 5.2, Example 2), we noted that polynomial functions are just finite linear combinations of products of id by itself plus a constant function, thus the above Proposition (i) shows continuity.

Rational functions are quotients of polynomial functions, defined where the denominator does not vanish, and are therefore also continuous by the second part of (i) in the above Proposition. □

EXAMPLE: 1) $p(x) := -x^2$ defines a continuous function on $\mathbb{R}$ and exp is continuous on $\mathbb{R}$. Hence the function $\exp \circ p\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \exp(-x^2)$ is continuous $\mathbb{R} \to \mathbb{R}$.

2) The *hyperbolic sine and cosine* ⟨*hyperbolischer Sinus und Cosinus*⟩ are defined by

$$\sinh(x) := \frac{\exp(x) - \exp(-x)}{2} \quad \text{and} \quad \cosh(x) := \frac{\exp(x) + \exp(-x)}{2} \qquad (x \in \mathbb{R}),$$

hence are continuous functions on $\mathbb{R}$.

5.7. Limit of a function:

Recall that $a \in \mathbb{R}$ is an adherent point of $D \subseteq \mathbb{R}$ if and only if there exists a sequence (x_n) in D (i.e., $x_n \in D$ for all n) such that $x_n \to a$ $(n \to \infty)$. If a is an element of D then the latter condition is clearly satisfied by the constant sequence $x_n = a$ for all n. In general, an adherent point of D need not be a member of the set D.

DEFINITION: Let $f\colon D \to \mathbb{R}$ and a an adherent point of D. The function f has *limit* $c \in \mathbb{R}$ *as* x *tends to* a, if every sequence (x_n) in D such that $x_n \to a$ $(n \to \infty)$ satisfies $\lim\limits_{n\to\infty} f(x_n) = c$. A short-hand notation for this fact is

$$\lim_{x\to a} f(x) = c \qquad \text{or} \qquad f(x) \to c \quad (x \to a).$$

We also define $c \in \mathbb{R}$ to be *the limit of f at a from the right* ⟨rechtsseitiger Grenzwert⟩

$$\lim_{x \searrow a} f(x) = c, \quad \text{or also} \quad \lim_{x \to a+} f(x) = c$$

if a is an adherent point of $D \cap [a, \infty[$ and for all sequences (x_n) with $x_n \in D$ and $x_n > a$ such that $x_n \to a$ we have $\lim f(x_n) = c$.
The notion of *limit from the left* ⟨linksseitiger Grenzwert⟩ $\lim\limits_{x \nearrow a} f(x)$, also denoted by $\lim\limits_{x \to a-} f(x)$, is defined analogously using $]-\infty, a] \cap D$ and $x_n < a$ instead.

Finally, we define *limits of f at infinity* as follows:

$$\lim_{x \to \infty} f(x) = c$$

means that D is unbounded from above and for every sequence (x_n) with $x_n \in D$ and $x_n \to \infty$ we have $\lim f(x_n) = c$.
We define $\lim\limits_{x \to -\infty} f(x)$ similarly when D is unbounded from below using $x_n \to -\infty$.

Of course, we will often find it convenient to also use the above notions with improper limits $c = \pm\infty$. The required adaptations of the definition should be routine and are left to the reader.

EXAMPLES: 1) For the rational function $f \colon \mathbb{R} \setminus \{1\} \to \mathbb{R}$, $f(x) = (x^2 - 1)/(x - 1)$, we have

$$\lim_{x \to 1} f(x) = 2.$$

Indeed, if $x_n \to 1$ with $x_n \neq 1$ then

$$f(x_n) = \frac{(x_n - 1)(x_n + 1)}{x_n - 1} = x_n + 1 \to 2 \quad (n \to \infty).$$

2) $\lim\limits_{x \searrow 1} \lfloor x \rfloor = 1$, since $\lfloor x_n \rfloor = 1$ when $1 < x_n < 2$. On the other hand, $\lim\limits_{x \nearrow 1} \lfloor x \rfloor = 0$ as $\lfloor x_n \rfloor = 0$ when $0 \leq x_n < 1$.
We conclude that $\lim\limits_{x \to 1} \lfloor x \rfloor$ does not exist, because otherwise the limits from the left and from the right would have to be equal.

3) Let $m \in \mathbb{N}$, $m \geq 1$, and $p \colon \mathbb{R} \to \mathbb{R}$ be a polynomial function of the form

$$p(x) = x^m + a_{m-1}x^{m-1} + \ldots + a_0.$$

Then we have $\lim\limits_{x \to \infty} p(x) = \infty$ and $\lim\limits_{x \to \infty} \dfrac{1}{p(x)} = 0.$

To see this, we first note that for all $x > 0$ we have the estimate

$$p(x) = x^m \left(1 + \frac{a_{m-1}}{x} + \ldots + \frac{a_0}{x^m}\right) \geq x^m \left(1 - \frac{|a_{m-1}|}{|x|} - \ldots - \frac{|a_0|}{|x^m|}\right).$$

Let $x \geq M := 2m \cdot \max(1, |a_{m-1}|, \dots, |a_0|)$, then the above inequality implies

$$p(x) \geq x^m \left(1 - m \cdot \frac{1}{2m}\right) = \frac{x^m}{2} \qquad \text{(in particular, } p(x) \geq \frac{1}{2}\text{)}.$$

Let $x_n \to \infty$ and choose $n_0 \in \mathbb{N}$ such that $x_n \geq M$ for all $n \geq n_0$. Then we obtain for $n \geq n_0$

$$p(x_n) \geq \frac{x_n^m}{2} \to \infty \qquad (n \to \infty),$$

therefore $\lim p(x_n) = \infty$, which proves the first assertion above. The second assertion follows immediately from the first, if we note that $1/p(x)$ is well-defined for $x \geq M$ (since $p(x) \geq 1/2$ then, as noted above).

Remark: Note that if $a \in D$ and $\lim_{x \to a} f(x)$ exists then the limit has to be $f(a)$ (since $x_n = a$ is a special sequence in D converging to a).

Warning: The notion of 'limit of a function' is not used in exactly the same way as we do here throughout the literature. Some texts (e.g. [Heu88]) require the admissible sequences (x_n) in the definition to be in $D \setminus \{a\}$, so that the special choice $x_n = a$ is excluded even in the case where a belongs to D. If a is an adherent point of D and does not belong to D, both notions give the same result concerning existence and value of the function limit. However, if $a \in D$ the conclusions may differ, as can be seen from the following example: Let $D := \mathbb{R} \setminus \{0\}$ and define $f \colon D \to \mathbb{R}$ by $f(x) := 1$ if $x \neq 0$, and $f(0) := 0$. Then in the sense of our definition f does not have a limit at 0, whereas we obtain for all sequences (x_n) with $x_n \neq 0$ and $x_n \to 0$ that $\lim f(x_n) = 1$ (note that this value differs from $f(0)$), which would give existence of the limit of f at 0 in the alternative definition.

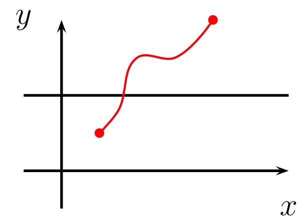
Since the notion of 'limit of a function' is essentially a short-hand notation to describe the way how a function translates converging sequences into sequences of corresponding function values, we can rephrase the sequence test of continuity 5.5 in these terms.

Proposition: A function $f \colon D \to \mathbb{R}$ is continuous at a point $a \in D$ if and only if

$$\lim_{x \to a} f(x) = f(a).$$

Proof. This is immediate from Theorem 5.5 and the remark made above. □

5.8. The intermediate value property ⟨*Zwischenwertsatz*⟩:



Theorem: Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous and $c \in \mathbb{R}$ lies between $f(a)$ and $f(b)$, that is $f(a) \leq c \leq f(b)$ or $f(b) \leq c \leq f(a)$. Then there exists $x_0 \in [a,b]$ such that $f(x_0) = c$.

In other words, a continuous function on $[a,b]$ attains every value between $f(a)$ and $f(b)$ at least once — there are no gaps in $f([a,b])$.

An important **special case** of the Theorem is the following: If $f\colon [a,b] \to \mathbb{R}$ is continuous and $f(a) < 0$ and $f(b) > 0$ (resp. $f(a) > 0$ and $f(b) < 0$), then f has a *zero* ⟨*Nullstelle*⟩ in $[a,b]$, i.e., $\exists\, x_0 \in [a,b]$: $f(x_0) = 0$.

Example of an Application: Let $p\colon \mathbb{R} \to \mathbb{R}$ be a polynomial function of odd degree $m = 2n+1$ (with $n \in \mathbb{N}$), say,

$$p(x) = b_{2n+1}x^{2n+1} + b_{2n}x^{2n} + \ldots + b_0 \qquad (x \in \mathbb{R}),$$

where $b_{2n+1} \neq 0$. Then p has at least one <u>real</u> zero.

To show this, we first write

$$p(x) = b_{2n+1} \cdot \big(x^{2n+1} + \frac{b_{2n}}{b_{2n+1}}x^{2n} + \ldots + \frac{b_0}{b_{2n+1}}\big) = b_{2n+1} \cdot q(x),$$

where the polynomial function q is of the form $q(x) = x^{2n+1} + a_{2n}x^{2n} + \ldots + a_0$ (with $a_j = b_j/b_{2n+1}$ for $j = 0, \ldots, 2n$).

By 5.7, Example 3), we have $\lim\limits_{x\to\infty} q(x) = \infty$, hence there exists $x_+ > 0$ such that $q(x_+) > 0$. Similarly, upon observing that

$$q(-x) = -x^{2n+1} + a_{2n}x^{2n} - \ldots + a_0 = -(x^{2n+1} - a_{2n}x^{2n} + \ldots - a_0)$$

we obtain that $\lim\limits_{x\to-\infty} q(x) = -\infty$, hence there exists $x_- < 0$ such that $q(x_-) < 0$.

Since $q\,|_{[x_-,x_+]}\colon [x_-,x_+] \to \mathbb{R}$ is continuous and $q(x_-) < 0 < q(x_+)$, the above Theorem implies that there exists $x_0 \in [x_-,x_+]$ such that $q(x_0) = 0$ (in fact, $x_- < x_0 < x_+$, because the values of q at $x_\pm$ are known to be nonzero). Therefore also $p(x_0) = b_{2n+1}\,q(x_0) = 0$.

Proof of the Theorem.

WLOG (:= without loss of generality) ⟨*OBdA (:= ohne Beschränkung der Allgemeinheit)*⟩ we may assume that $f(a) < c < f(b)$. [otherwise we just have to consider $-f$ instead]

If $c \neq 0$ we can reduce the statement to that of the special case of a zero by putting $f_1(x) := f(x) - c$. Then $f_1(a) = f(a) - c < 0$ and $f_1(b) = f(b) - c > 0$ and the assertion of the Theorem is equivalent to the existence of a zero $x_0 \in [a,b]$ of the function f_1.

So, again WLOG we may assume that $c = 0$ and $f(a) < 0 < f(b)$. We have to show that there is some $x_0 \in [a,b]$ such that $f(x_0) = 0$.

We will find x_0 by constructing a sequence of nested intervals in the fashion of a so-called *bisection method*. To be more precise, we claim that we can define $[a_n, b_n] \subseteq [a,b]$ for all $n \in \mathbb{N}$ with the following properties:

1. $\forall n \in \mathbb{N}$: $[a_n, b_n] \subseteq [a_{n-1}, b_{n-1}]$
2. $b_n - a_n = \dfrac{b-a}{2^n}$
3. $f(a_n) < 0$ and $f(b_n) \geq 0$.

Put $a_0 := a$ and $b_0 := b$, then properties 2 and 3 are satisfied. We proceed inductively and assume that $[a_0, b_0]$, ..., $[a_n, b_n]$ have been defined satisfying properties 1-3. Let $m := (b_n - a_n)/2$ (this is the midpoint of $[a_n, b_n]$) and distinguish two cases:

If $f(m) \geq 0$ put $a_{n+1} := a_n$ and $b_{n+1} := m$

If $f(m) < 0$ put $a_{n+1} := m$ and $b_{n+1} := b_n$.

Then properties 1–3 are valid for $[a_{n+1}, b_{n+1}]$ as well. By the principle of nested intervals we obtain that (a_n) and (b_n) converge to the same limit $x_0 \in [a,b]$, that is $\lim a_n = \lim b_n = x_0$. Since f is continuous we have that

$$f(x_0) = \lim f(a_n) = \lim f(b_n).$$

By property 3 we obtain in addition

$$f(x_0) = \lim f(a_n) \leq 0 \leq \lim f(b_n) = f(x_0),$$

which proves that $f(x_0) = 0$. □

Corollary: Let $I \subseteq \mathbb{R}$ be a nonempty interval and $f\colon I \to \mathbb{R}$ be continuous. Then $f(I) \subseteq \mathbb{R}$ is an interval as well.

Proof. Let $A := \inf f(I)$ and $B := f(I)$, where we allow for the improper values $A = -\infty$ (unbounded below) and $B = \infty$ (unbounded above). If $A = B$ then $f(I)$ contains just a single point, in which case the statement is true. So we henceforth assume that $A < B$.

We assert that $]A, B[\subseteq f(I)$: Let $y \in]A, B[$, then there exist $r, s \in I$ such that $f(r) < y < f(s)$. By the above Theorem we have some $x_0 \in I$ such $f(x_0) = y$. Therefore $y \in f(I)$.

To summarize, $]A, B[\subseteq f(I) \subseteq [A, B]$, hence $f(I)$ equals one of the intervals $]A, B[$ or $]A, B]$ or $[A, B[$ or $[A, B]$. □

5.9. Continuous functions on bounded closed intervals:

DEFINITION: A function $f\colon D \to \mathbb{R}$ is called *bounded* ⟨*beschränkt*⟩ if the image set $f(D) \subseteq \mathbb{R}$ is bounded, i.e.,

$$\exists M > 0 \;\; \forall x \in D: \;\; |f(x)| \leq M.$$

THEOREM: Let $f\colon [a,b] \to \mathbb{R}$ be continuous. Then f is bounded and attains maximum and minimum values, i.e., there exist $x_1, x_2 \in [a,b]$ such that

$$f(x_1) = \min f([a,b]) = \min \{f(x) : x \in [a,b]\} \qquad (= \inf f([a,b]))$$
$$f(x_2) = \max f([a,b]) = \max \{f(x) : x \in [a,b]\} \qquad (= \sup f([a,b])).$$

REMARK: (i) In the hypothesis of this theorem it is essential that the interval $[a,b]$, where f is defined and continuous, is <u>bounded</u> (i.e., $-\infty < a \leq b < \infty$) and <u>closed</u> (i.e., the boundary points a and b belong to the interval). Otherwise the statement is not true in general as can be seen from the following examples: Consider the continuous functions

$$f_1\colon]0,1] \to \mathbb{R}, x \mapsto \frac{1}{x}, \qquad f_2\colon]0,1[\to \mathbb{R}, x \mapsto x, \qquad f_3\colon [0,\infty[\to \mathbb{R}, x \mapsto x.$$

Then f_1 and f_3 are unbounded and do not attain a maximum, f_2 does neither attain a maximum nor a minimum.

(ii) As is shown by the simple example of a constant function, the locations x_1 and x_2 of a minimum or a maximum need not be unique.

Proof of the Theorem. It suffices to give the proof for boundedness from above and concerning the maximum, the case of minimum and boundedness from below can be reduced to the latter by switching to $-f$.

Let $A := \sup f([a,b]) \in \mathbb{R} \cup \{\infty\}$, then there exists a sequence (a_n) in $[a,b]$ such that $f(a_n) \to A$ $(n \to \infty)$.

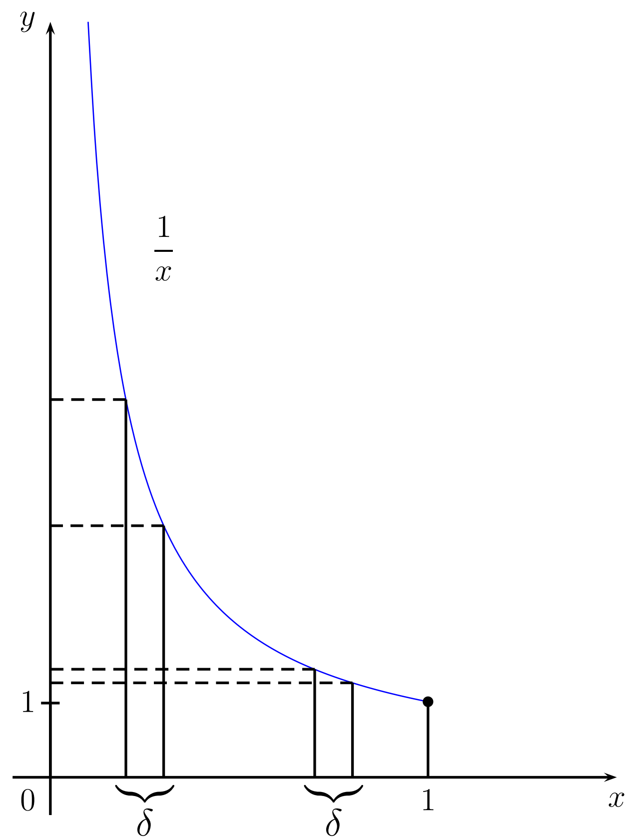
Since $[a,b]$ is a bounded subset of $\mathbb{R}$ the Theorem of Bolzano-Weierstraß implies that there is a convergent subsequence $(a_{n_k})_{k\in\mathbb{N}}$. Let $x_2 := \lim\limits_{k\to\infty} a_{n_k} \in [a,b]$.

Since f is continuous we obtain that

$$\mathbb{R} \ni f(x_2) = \lim_{k\to\infty} f(a_{n_k}) = A = \sup f([a,b]).$$

In particular, $f([a,b])$ is bounded above and the supremum is a maximum, which is attained by f at $x_2 \in [a,b]$. □

5.10. Uniform continuity: If we are to check continuity of a real valued function f at a certain point x_0 in its domain D, then for given $\varepsilon > 0$ we have to find $\delta > 0$ such that the condition $|f(x) - f(x_0)| < \varepsilon$ is met whenever $x \in D$ satisfies $|x - x_0| < \delta$. We observe that, in general, δ will dependend on $\varepsilon > 0$ as well as on the point x_0. Consider the following example, where the range of possible values for δ is strictly shrinking as ε gets smaller or x_0 varies: Let $D =]0,1]$ and $f\colon D \to \mathbb{R}$ with $f(x) = 1/x$, which is continuous in every point $x_0 \in D$.



Fix some $x_0 \in D$ and $\varepsilon > 0$ arbitrarily and let us test the allowed tolerance in varying the argument in $0 < x \leq x_0$ while maintaining $|f(x) - f(x_0)| < \varepsilon$. For every $0 < \delta < x_0$ let $x_\delta := x_0 - \delta$. Then

$$|f(x_\delta) - f(x_0)| = \frac{1}{x_\delta} - \frac{1}{x_0} = \frac{x_0 - x_\delta}{x_0\, x_\delta} = \frac{\delta}{x_0\,(x_0 - \delta)}.$$

Thus requiring $|f(x) - f(x_0)| < \varepsilon$ for all $x \in]0,1]$ with $|x - x_0| < \delta$ implies $\varepsilon > \frac{\delta}{x_0^2 - x_0\delta}$. Equivalently, $\varepsilon x_0^2 - \varepsilon x_0 \delta > \delta$ and hence

$$\delta < \frac{\varepsilon x_0^2}{1 + \varepsilon\, x_0} < \varepsilon\, x_0^2.$$

This shows that the smaller $x_0 > 0$ is the smaller we have to choose $\delta > 0$ (even at fixed value of $\varepsilon > 0$).

We thus obtain a stronger form of continuity notion, if we require that for each $\varepsilon > 0$ a suitable $\delta > 0$ can be found which guarantees the typical ε-δ-estimate to hold for all pairs of points in the domain of relative distance less than δ.

Definition: Let $D \subseteq \mathbb{R}$. A function $f\colon D \to \mathbb{R}$ is *uniformly continuous* ⟨*gleichmäßig stetig*⟩ (in D), if the following holds:

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, x' \in D:\ |x - x'| < \delta \Longrightarrow |f(x) - f(x')| < \varepsilon.$$

Remark: It is immediate from the definition that every uniformly continuous function is continuous (at every point in the domain). The converse is not true as is illustrated by the example above with $D =]0,1]$, $f(x) = 1/x$: If $x_n = 1/n$ and $x'_n = 1/(2n)$ ($n \in \mathbb{N}$, $n \geq 1$) then

$$|x_n - x'_n| = \frac{1}{n} - \frac{1}{2n} = \frac{1}{2n}$$

is arbitrarily small when n is sufficiently large, but

$$|f(x_n) - f(x'_n)| = 2n - n = n$$

is unbounded, hence will never stay below a given ε-tolerance.

However, as the following theorem will show, there is no distinction between continuity and uniform continuity if D is a bounded closed interval.

THEOREM: If $f\colon [a,b] \to \mathbb{R}$ is continuous then f is uniformly continuous (on $[a,b]$).

Proof. (by contradiction) If f is not uniformly continuous then

$$\exists \varepsilon > 0\ \forall n \in \mathbb{N}, n > 0\ \exists x_n, x'_n \in [a,b]:\ |x_n - x'_n| < \frac{1}{n} \text{ and } |f(x_n) - f(x'_n)| \geq \varepsilon.$$

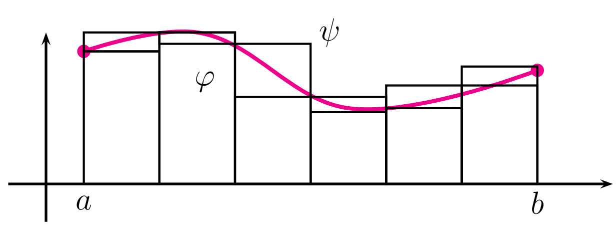
The sequence (x_n) is bounded, thus by the Theorem of Bolzano-Weierstraß possesses a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$. Let $x_0 := \lim x_{n_k} \in [a,b]$.

Since $|x_{n_k} - x'_{n_k}| < 1/n_{n_k} \to 0$ $(k \to \infty)$ we have that $\lim x'_{n_k} = \lim x_{n_k} = x_0$. Then the continuity of f at x_0 yields

$$0 < \varepsilon \leq |f(x_{n_k}) - f(x'_{n_k})| \to 0 \qquad (k \to \infty),$$

— a contradiction ↯ . □

5.11. Approximation by step (or simple) functions: As an application of the above Theorem 5.10 we show that "the area under the graph of a continuous function" can be approximated by sums over areas of small vertical rectangles. This will be used later in the chapter on integration theory.



The rectangles can be represented as graphs of step or simple functions and the approximation result is stated in terms of these as a uniform approximation from above and from below.

PROPOSITION: Let $f\colon [a,b] \to \mathbb{R}$ be continuous. For every $\varepsilon > 0$ there exist simple functions $\varphi, \psi\colon [a,b] \to \mathbb{R}$ with the following properties valid for all $x \in [a,b]$:

(a) $\varphi(x) \le f(x) \le \psi(x)$

(b) $|\varphi(x) - \psi(x)| = \varphi(x) - \psi(x) \le \varepsilon$.

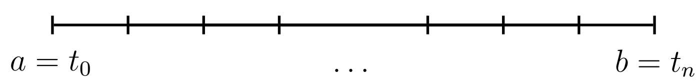
Proof. By Theorem 5.10 f is uniformly continuous on $[a,b]$. Therefore we can find $\delta > 0$ such that

$$\forall x, x' \in [a,b] : \ |x - x'| < \delta \Longrightarrow |f(x) - f(x')| < \varepsilon.$$

Choose $n \in \mathbb{N}$ large enough to ensure $(b-a)/n < \delta$ and define partition points

$$t_k := a + k \cdot \frac{b-a}{n} \qquad (k = 0, \ldots, n).$$

In this way we obtain an equidistant partition of $[a,b]$



by $t_0 = a < t_1 < \ldots < t_n = b$ with

$$t_k - t_{k-1} = \frac{b-a}{n} < \delta.$$

As heights of the approximating rectangles we choose the maximum or minimum values of f on the corresponding subintervals $[t_{k-1}, t_k]$ $(k = 1, \ldots, n)$ of the partition, that is we define

$$c_k := \max\{f(x) : t_{k-1} \le x \le t_k\}, \quad c'_k := \min\{f(x) : t_{k-1} \le x \le t_k\}.$$

By Theorem 5.9 there exist $\xi_k, \xi'_k \in [t_{k-1}, t_k]$ such that $f(\xi_k) = c_k$ and $f(\xi'_k) = c'_k$ $(k = 1, \ldots, n)$. Since $|\xi_k - \xi'_k| < \delta$ we have $|c_k - c'_k| < \varepsilon$ from the uniform continuity property noted in the beginning.

Finally, we define the simple functions $\varphi, \psi\colon [a,b] \to \mathbb{R}$ as follows:

Let $\varphi(a) := f(a)$ and $\psi(a) := f(a)$,

for $t_{k-1} < x \le t_k$ we set $\varphi(x) := c_k$ and $\psi(x) := c'_k$ $(k = 1, \ldots, n)$.

Then the conditions (a) and (b) follow by construction of φ and ψ. □

5.12. Continuous inverse function: Let $A, B \subseteq \mathbb{R}$. Assume that the function $f\colon A \to B$ is bijective, then the inverse function $f^{-1}\colon B \to A$ exists. If we know that f is continuous, does this imply that f^{-1} is also continuous? In general, the answer is 'no' (see the exercises for an example).

It turns out that there is a positive answer to the above question under the two additional hypotheses of strict monotonicity on f and that A is an interval.

Recall that f is strictly increasing (resp. decreasing) if $x_1 < x_2$ implies $f(x_1) < f(x_2)$ (resp. $f(x_1) > f(x_2)$) and that a strictly monotone function necessarily is injective.

THEOREM: Let $I \subseteq \mathbb{R}$ be an interval and $f\colon I \to \mathbb{R}$ be continuous and strictly increasing (resp. decreasing). Then f maps the interval I bijectively onto the interval $J := f(I)$ and the corresponding inverse function $J \to I$ is also continuous and strictly increasing (resp. decreasing).

(Strictly speaking, we deal here with the inverse of the map $\tilde{f}\colon I \to J$, $x \mapsto f(x)$; but we will follow the common abuse of language and denote $\tilde{f}$ again by f and its inverse by $f^{-1}\colon J \to I$.)

Proof. We present the proof for the case that f is strictly increasing, the case where f is strictly decreasing is reduced to this by considering $-f$ instead.

Corollary 5.8 implies that $J = f(I)$ is an interval. Since f is strictly increasing it is injective, hence f is bijective as a map $I \to J$. Let $f^{-1}\colon J \to I$ denote the inverse of this map.

Note that for $x_1, x_2 \in I$ the inequality $f(x_1) < f(x_2)$ in turn implies $x_1 < x_2$ (since then $x_1 = x_2$ is impossible with different function values and $x_1 > x_2$ contradicts the fact that f increases), therefore we have

$$\forall x_1, x_2 \in I: \quad x_1 < x_2 \Longleftrightarrow f(x_1) < f(x_2),$$

which shows that f^{-1} is strictly increasing as well.

It remains to prove that f^{-1} is continuous at every point $b \in J$.

Case 1, if $b \in J$ is not a boundary point of J: Let $a := f^{-1}(b) \in I$. Then a is not a boundary point of I (for otherwise by monotonicity b would have to be boundary point of J). Choose $\varepsilon > 0$ so small that both $a - \varepsilon$ and $a + \varepsilon$ belong to I. Since f is strictly increasing we have

$$f(a - \varepsilon) < f(a) = b < f(a + \varepsilon).$$

Thus we can find $\delta > 0$ such that

$$f(a - \varepsilon) < b - \delta < b + \delta < f(a + \varepsilon),$$

which simply means that

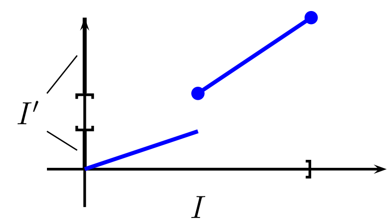
$$f^{-1}(U_\delta(b)) \subseteq U_\varepsilon(f^{-1}(b))$$

and therefore proves the continuity of f^{-1} at b.

Case 2, if $b \in J$ is the left boundary point of J: Then $a := f^{-1}(b)$ has to be the left boundary point of I (since f is strictly increasing). We can copy the proof of case 1 with the only changes that we use $U_\delta \cap J$ and $U_\varepsilon(f^{-1}(b)) \cap I$ as neighborhoods and the chain of inequalities reads $f(a) = b < b + \delta < f(a + \varepsilon)$.

Case 3, if $b \in J$ is the right boundary point of J: Similarly to case 2. □

REMARK: The second part of the above proof shows, in fact, the following result: If $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is strictly increasing (not necessarily continuous!), then $f^{-1} : f(I) \to I$ is continuous. But $f(I)$ need not be an interval, if f is discontinuous:



ROOT FUNCTIONS: As an application of the above Theorem we consider for any $k \in \mathbb{N}$, $k \geq 1$, the functions[4]

$$f_{2k} \colon [0, \infty[\to [0, \infty[,\ x \mapsto x^{2k}, \qquad \text{and} \qquad f_{2k+1} \colon \mathbb{R} \to \mathbb{R},\ x \mapsto x^{2k+1}.$$

All these functions are continuous, strictly increasing, and bijective, therefore the corresponding inverse functions

$$f_{2k}^{-1} \colon [0, \infty[\to [0, \infty[\qquad \text{and} \qquad f_{2k+1}^{-1} \colon \mathbb{R} \to \mathbb{R}$$

are continuous and strictly increasing. We use the the following notation for their function values (for x in the appropriate domain)

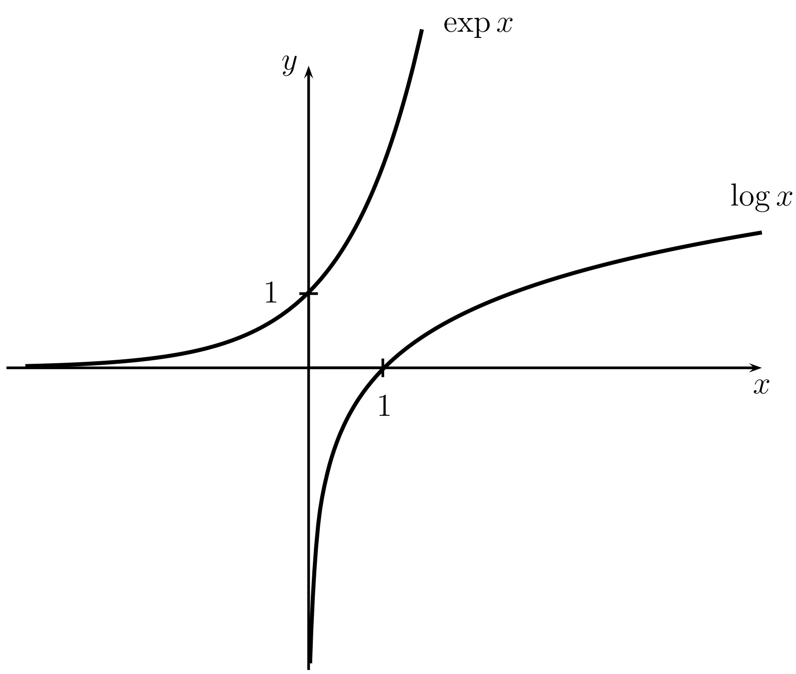
$$n \geq 2 : \quad \sqrt[n]{x} = x^{\frac{1}{n}} := f_n^{-1}(x).$$

[4]Altough the origin of the radical symbol $\sqrt{\ }$ is rather unclear, many believe that it is an abbreviation of the Latin word radix (root). The symbol was first used in Germany in the 16th century without the winkulum (i.e. the term $\sqrt{a+b}$ was originally denoted by $\sqrt{}(a+b)$)

§6. ELEMENTARY TRANSCENDENTAL FUNCTIONS

6.1. Proposition: The exponential function $\exp\colon \mathbb{R} \to \mathbb{R}$ is continuous, strictly increasing, and $\exp(\mathbb{R}) =]0,\infty[$. Its inverse function $\log\colon]0,\infty[\to \mathbb{R}$ is continuous, strictly increasing and is called the *natural logarithm* ⟨natürlicher Logarithmus⟩.[1] Furthermore, the following functional equation holds for all $x, y \in]0,\infty[$:

(6.1) $$\log(x \cdot y) = \log(x) + \log(y).$$



[1]Logarithms have been introduced by the Scottish mathematician John Napier in 1614, the term logarithm being derived from the Greek words λόγος ("proportion") and ἀριθμός ("number").
Nowadays, there are different notations used for logarithmic functions: While mathematicians often write $\log(x)$ for the natural logarithm and $\log_b(x)$ for the base-b logarithm, in many calculus textbooks a notations such as $\ln(x)$ can be found for the natural logarithm, $\lg(x)$ for the base-10 logarithm etc. In these lecture notes we will always use $\log(x)$ to denote the natural logarithm and $\log_b(x)$ for the base-b logarithm.

Proof. The continuity of exp has already been established in the previous section.

Step 1: We show that exp is strictly increasing.

For every $\xi > 0$ we have

$$\exp(\xi) = 1 + \xi + \sum_{k=2}^{\infty} \frac{\xi^k}{k!} > 1 + \xi > 1.$$

Let $x_1 < x_2$ then $\xi := x_2 - x_1 > 0$ and

$$\exp(x_2) = \exp(x_1 + \xi) = \exp(x_1) \cdot \exp(\xi) > \exp(x_1).$$

Step 2: We show that $\exp(\mathbb{R}) =]0, \infty[$.

Since $\exp(x) > 0$ for all $x \in \mathbb{R}$ [4.23] we have $\exp(\mathbb{R}) \subseteq]0, \infty[$. To show the reverse inclusion relation, it suffices to show that

$$\lim_{n\to\infty} \exp(n) = \infty \quad \text{ and } \quad \lim_{n\to\infty} \exp(-n) = 0,$$

since then by the intermediate value theorem all values in $]0, \infty[$ are indeed attained.

For $n \in \mathbb{N}$ we had shown $\exp(n) = e^n$. Since $e > 2$ we therefore have $e^n \to \infty$ $(n \to \infty)$, which implies that

$$\exp(-n) = \frac{1}{\exp(n)} = \frac{1}{e^n} \to 0 \quad (n \to \infty).$$

We may thus define the function $f \colon \mathbb{R} \to]0, \infty[$, $f(x) := \exp(x)$, which is again continuous and strictly increasing. Due to Theorem 5.12 the inverse function $\log := f^{-1} \colon]0, \infty[\to \mathbb{R}$ is also continuous and strictly increasing.

Step 3: We prove the functional equation (6.1).

Let $x, y \in]0, \infty[$ and put $\xi := \log(x)$, $\eta := \log(y)$. Then $\exp(\xi + \eta) = \exp(\xi) \cdot \exp(\eta) = x \cdot y$ and therefore

$$\log(x \cdot y) = \xi + \eta = \log(x) + \log(y).$$

□

As a simple consequence we obtain

$$\log(x^k) = k \log(x) \qquad \text{for all } x > 0 \text{ and } k \in \mathbb{N}.$$

6.2. Real powers and general exponentials: We can use the exponential function and the logarithm to give a simple definition of expressions of the form r^s when $r > 0$ and $s \in \mathbb{R}$. Observe that if s is a natural number then $r^s = \exp(\log(r^s)) = \exp(s \log(r))$.

Definition: (i) Let $r > 0$ and $s \in \mathbb{R}$ then

$$r^s := \exp(s\log(r)) \in \,]0, \infty[.$$

As an immediate consequence of this definition we thus obtain the formula

$$\log(r^s) = s\log(r) \qquad (r > 0, s \in \mathbb{R}).$$

(ii) For any $\alpha \in \mathbb{R}$ we define general power or root functions $w_\alpha \colon]0, \infty[\to \mathbb{R}$ by

$$x \mapsto x^\alpha = \exp(\alpha \log(x)).$$

(iii) The Exponential function with base $a \in]0, \infty[$ is given by

$$\exp_a \colon \mathbb{R} \to \mathbb{R}, \quad \exp_a(x) := a^x = \exp(x\log(a)).$$

Note that $\exp(x) = \exp_e(x) = e^x$ for all $x \in \mathbb{R}$.

We list the basic properties of exponential function with base $a > 0$, which are immediate consequences of those for the exponential function and the logarithm.

Proposition: $\exp_a$ is continuous on $\mathbb{R}$ and we have the following:

(i) If $a > 1$ then $\exp_a$ is strictly increasing, if $0 < a < 1$ then $\exp_a$ is strictly decreasing.

(ii) The functional equation: $a^{x+y} = a^x \cdot a^y$ for all $x, y \in \mathbb{R}$.

(iii) Let $a > 0$. For all $m \in \mathbb{Z}$: $\exp_a(m) = a^m = a \cdot a \cdots a$ (m factors).
(In other words, the notation a^m is consistent with the algebraically defined integer powers.)

(iv) Let $a > 0$. If $p \in \mathbb{Z}$ and $q \in \mathbb{N}$, $q \geq 1$, then $a^{\frac{p}{q}} = \sqrt[q]{a^p} = (a^p)^{\frac{1}{q}}$.
(Consistency with the root functions as defined in 5.12.)

(v) Let $a > 0$. For all $x, y \in \mathbb{R}$: $(a^x)^y = a^{xy} = (a^y)^x$.

(vi) For all $a > 0$, $b > 0$, and $x \in \mathbb{R}$: $a^x \cdot b^x = (a \cdot b)^x$.

(vii) Let $a > 0$. For all $x \in \mathbb{R}$: $(\frac{1}{a})^x = a^{-x}$.

Proof. Immediate from the definition. □

6.3. A collection of useful limits:

1) For all $k \in \mathbb{N}$: $\lim\limits_{x\to\infty} \frac{e^x}{x^k} = \infty$.

We may assume that $x > 0$, which yields $e^x = \sum\limits_{n=0}^{\infty} \frac{x^n}{n!} > \frac{x^{k+1}}{(k+1)!}$. Therefore

$$\frac{e^x}{x^k} > \frac{x}{(k+1)!} \to \infty \qquad \text{as } x \to \infty.$$

2) For all $k \in \mathbb{N}$: $\lim_{x\to\infty} \frac{x^k}{e^x} = 0$. (Follows directly from 1).)

3) For all $k \in \mathbb{N}$: $\lim_{x\searrow 0} x^k e^{1/x} = \infty$.

Writing $y = 1/x$ gives $\lim_{x\searrow 0} x^k e^{1/x} = \lim_{y\to\infty} \frac{e^y}{y^k} = \infty$ (by 1)).

4) $\lim_{x\to\infty} \log(x) = \infty$ and $\lim_{x\searrow 0} \log(x) = -\infty$.

Both assertions follows from Proposition 6.1, which implies that $\log\colon]0,\infty[\to \mathbb{R}$ is strictly increasing and bijective.

5) For all $\alpha > 0$: $\lim_{x\searrow 0} x^\alpha = 0$ and $\lim_{x\searrow 0} x^{-\alpha} = \infty$.

The second assertion follows from the first. To prove the first we write $x = e^{-y/\alpha}$ (equivalently, $y = -\alpha \log(x)$) and compute

$$\lim_{x\searrow 0} x^\alpha = \lim_{y\to\infty} e^{-y} = 0.$$

6) For all $\alpha > 0$: $\lim_{x\to\infty} \frac{\log(x)}{x^\alpha} = 0$.

We may assume that $x > 0$ and write $x^\alpha = e^y$ (equivalently, $y = \alpha \log(x)$) to obtain

$$\lim_{x\to\infty} \frac{\log(x)}{x^\alpha} = \frac{1}{\alpha} \cdot \lim_{y\to\infty} \frac{y}{e^y} = 0.$$

7) For all $\alpha > 0$: $\lim_{x\searrow 0} x^\alpha \log(x) = 0$.

Upon writing $x = 1/y$ (so that $y \to \infty$) we have $x^\alpha \log(x) = -\log(y)/y^\alpha$, then use 6).

8) $\lim_{\substack{x\to 0 \\ x\neq 0}} \frac{e^x - 1}{x} = 1$

The remainder term estimate (4.4), (4.5) for the exponential sum gives for all x with $|x| \le 3/2$ that

$$|e^x - (1+x)| \le 2\,\frac{|x|^2}{2!} = |x|^2.$$

In other words, if $0 < |x| \le 3/2$ then

$$\left|\frac{e^x - 1}{x} - 1\right| = \frac{|e^x - 1 - x|}{|x|} \le |x| \to 0 \quad \text{as } x \to 0.$$

6.4. The Landau-symbols[2] — comparison of asymptotic growth:

DEFINITION: (i) Let $a \in \mathbb{R}$ and $f, g\colon]a, \infty[\to \mathbb{R}$. We write

$$f(x) = o(g(x)) \quad (x \to \infty),$$

to mean that $\forall \varepsilon\ \exists R > a$: $|f(x)| \leq \varepsilon\, |g(x)|$ holds for all $x \geq R$.
("$f(x)$ is a little-oh of $g(x)$ as x tends to infinity".)

We write

$$f(x) = O(g(x)) \quad (x \to \infty),$$

to mean that $\exists K > 0\ \exists R > a$: $|f(x)| \leq K\, |g(x)|$ holds for all $x \geq R$.
("$f(x)$ is a big-oh of $g(x)$ as x tends to infinity".)

(ii) Let $D \subseteq \mathbb{R}$ and x_0 be an adherent point of D. We write

$$f(x) = o(g(x)) \quad (x \to x_0, x \in D),$$

to mean that $\forall \varepsilon > 0\ \exists \delta > 0$: $|f(x)| \leq \varepsilon\, |g(x)|$ holds $\forall x \in U_\delta(x_0) \cap D$.

We write

$$f(x) = O(g(x)) \quad (x \to x_0, x \in D),$$

to mean that $\exists K > 0\ \exists \delta > 0$: $|f(x)| \leq K\, |g(x)|$ holds $\forall x \in U_\delta(x_0) \cap D$.

REMARK: (i) If, for example, $g(x) \neq 0$ for all x near x_0 and $\lim\limits_{x \to x_0} \dfrac{f(x)}{g(x)} = 0$, then $f(x) = o(g(x))$ $(x \to x_0)$.

(ii) We will occasionally make use of a notation like

$$f(x) = h(x) + O(g(x))$$

to mean that $f(x) - h(x) = O(g(x))$.

EXAMPLES: 1) If $\alpha > 0$ then $\log(x) = o(x^\alpha)$ $(x \to \infty)$ [cf. 6.3.6)].

2) $e^x = 1 + x + O(x^2)$ $(x \to 0)$, since $|e^x - 1 - x| \leq |x|^2$ when $|x| \leq 3/2$ [as seen in 6.3.8)].

3) $f(x) = f(x_0) + o(1)$ $(x \to x_0)$ $\iff$ $\lim\limits_{x \to x_0} f(x) = f(x_0)$ $\iff$ f is continuous at x_0.

4) If p is a polynomial function of degree m, then $p(x) = O(x^m)$ $(x \to \infty)$.

---

[2] Edmund Georg Hermann Landau (1877–1938) [ˈɛdmʊnt ˈgeˑɔʁk ˈhɛʁman ˈlandaʊ], German mathematician

6.5. A digression into basic analysis on $\mathbb{C}$:

Let $z = x + iy \in \mathbb{C}$, so that $x = \mathrm{Re}(z)$, $y = \mathrm{Im}(z)$ and $\overline{z} = x + iy \in \mathbb{C}$. Then the product $z\overline{z} = (x+iy)(x-iy) = x^2+y^2$ always gives a non-negative real number and we may define the *absolute value* of z by

$$|z| := \sqrt{z\overline{z}} = \sqrt{x^2+y^2} = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}.$$

Note that, since real numbers x are embedded as complex numbers[3] of the form $x + i0$, the absolute value of x as a real number is the same as its absolute value as a complex number.

LEMMA: The absolute value as a map $|.|\colon \mathbb{C} \to \mathbb{R}$ has the following properties, valid for all $z, z_1, z_2 \in \mathbb{C}$:

(i) $|z| \geq 0$ and $|z| = 0 \Leftrightarrow z = 0$

(ii) $|\overline{z}| = |z|$

(iii) $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$

(iv) $|\mathrm{Re}(z)| \leq |z|$ and $|\mathrm{Im}(z)| \leq |z|$

(v) $|z_1 + z_2| \leq |z_1| + |z_2|$ (triangle inequality).

Proof. Let $z = x + iy$, $z_k = x_k + iy_k$ $(k = 1, 2)$.

(i): $|z| \geq 0$ and $|0| = 0$ is immediate. If $|z| = 0$, then $0 \leq x^2 \leq x^2 + y^2 = 0$ as well as $0 \leq y^2 \leq x^2 + y^2 = 0$, hence $x = 0$ and $y = 0$.

(ii): Clear from the definition.

(iii): $|z_1z_2|^2 = (z_1z_2)(\overline{z_1z_2}) = (z_1\overline{z_1})(z_2\overline{z_2}) = |z_1|^2|z_2|^2$.

(iv): $|x|^2 = x^2 \leq x^2 + y^2$ and $|y|^2 = y^2 \leq x^2 + y^2$.

(v): $|z_1 + z_2|^2 = (z_1 + z_2)(\overline{z_1} + \overline{z_2}) = |z_1|^2 + 2\,\mathrm{Re}(z_1\overline{z_2}) + |z_2|^2 \leq$ [by (iv)]
$|z_1|^2 + 2|z_1||z_2| + |z_2|^2 = (|z_1| + |z_2|)^2$. □

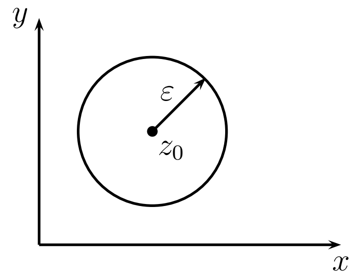
(a) Convergence in $\mathbb{C}$:

DEFINITION: (i) Let $z_0 \in \mathbb{C}$ and $\varepsilon > 0$, then the *ε-neighborhood* $U_\varepsilon(z_0)$ of z_0 is defined by

$$U_\varepsilon(z_0) := \{z \in \mathbb{C} : |z - z_0| \leq \varepsilon\}.$$

In a planar representation of the complex numbers, $U_\varepsilon(z_0)$ is an open disk with radius ε around z_0:

[3] Square roots of negative numbers were "invented" by the Italian mathematicians Gerolamo Cardano and Raffaele Bombelli. In modern mathematics, complex numbers are generally denoted by $z = a + bi$ or sometimes $z = a + bj$.



(ii) A *sequence of complex numbers* is a map $c\colon \mathbb{N} \to \mathbb{C}$. We use the notation $(c_n)_{n\in\mathbb{N}}$ with $c_n := c(n)$.

(iii) The complex sequence (c_n) *converges to* $z_0 \in \mathbb{C}$, denoted by $c_n \to z_0$ $(n \to \infty)$ or $\lim\limits_{n\to\infty} c_n = z_0$, if

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n : \quad |c_n - z_0| < \varepsilon.$$

Equivalently, we may require that

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n : \quad c_n \in U_\varepsilon(z_0).$$

Proposition: Let (c_n) be a complex sequence. Then the following are equivalent:

(i) (c_n) is convergent (in $\mathbb{C}$).

(ii) Both sequences $(\operatorname{Re} c_n)$ and $(\operatorname{Im} c_n)$ converge (in $\mathbb{R}$).

In this case we have $\lim c_n = \lim \operatorname{Re} c_n + i \lim \operatorname{Im} c_n$.

Proof. Let $a_n := \operatorname{Re} c_n$, $b_n := \operatorname{Im} c_n$ $(n \in \mathbb{N})$.

(i) ⇒ (ii): Let $c := \lim c_n$, $a := \operatorname{Re} c$ and $b := \operatorname{Im} c$.

If $\varepsilon > 0$ is given arbitrarily, we can find $n_0 \in \mathbb{N}$ such that $|c_n - c| < \varepsilon$ holds for all $n \geq n_0$. Therefore we have for all $n \geq n_0$

$$|a_n - a| = |\operatorname{Re}(c_n - c)| \leq |c_n - c| < \varepsilon \quad \text{as well as} \quad |b_n - b| = |\operatorname{Re}(c_n - c)| \leq |c_n - c| < \varepsilon,$$

which proves that $a_n \to a$ and $b_n \to b$.

(ii) ⇒ (i): Let $\varepsilon > 0$. Put $a := \lim a_n$, $b := \lim b_n$, and $c := a + ib$. Choose $n_0 \in \mathbb{N}$ such that $|a_n - a| < \varepsilon/2$ and $|b_n - b| < \varepsilon/2$ holds for all $n \geq n_0$. Then we have for $n \geq n_0$

$$|(a_n + ib_n) - (a + ib)| = |(a_n - a) + i(b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

thus $c_n \to c$ $(n \to \infty)$.

□

COROLLARY: If (c_n) is a convergent complex sequence, then $\lim \overline{c_n} = \overline{\lim c_n}$.

Proof. $\overline{\lim c_n} = \lim \operatorname{Re} c_n - i \operatorname{Im} c_n = \lim \operatorname{Re} \overline{c_n} + i \lim \operatorname{Im} \overline{c_n} = \lim \overline{c_n}$. □

Precisely as in the case of real sequences one proves the following rules for basic operations with convergent sequences:

If (c_n), (d_n) are convergent complex sequences and $\lambda \in \mathbb{C}$, then

$$\begin{aligned}
\lim(c_n + d_n) &= \lim c_n + \lim d_n \\
\lim(\lambda c_n) &= \lambda \lim c_n \\
\lim(c_n d_n) &= (\lim c_n)(\lim d_n) \\
\lim \frac{c_n}{d_n} &= \frac{\lim c_n}{\lim d_n} \quad \text{(if } d_n \neq 0 \text{ for almost all } n\text{).}
\end{aligned}$$

THEOREM (COMPLETENESS OF $\mathbb{C}$): A sequence (c_n) of complex numbers converges if and only if it is a *Cauchy sequence*, i.e.,

$$\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N} \; \forall n, m \geq n_0 : \quad |c_n - c_m| < \varepsilon. \tag{6.2}$$

Proof. (6.2) $\Leftrightarrow$ both $(\operatorname{Re} c_n)$ and $(\operatorname{Im} c_n)$ are Cauchy sequences in $\mathbb{R}$ $\Leftrightarrow$ both $(\operatorname{Re} c_n)$ and $(\operatorname{Im} c_n)$ are convergent in $\mathbb{R}$ $\underset{\text{[Prop.]}}{\Leftrightarrow}$ (c_n) is convergent in $\mathbb{C}$. □

(b) Complex series:

DEFINITION: Let (c_n) be a sequence of complex numbers. The series $\sum_{k=0}^{\infty} c_k$ is *convergent*, if the corresponding sequence (s_n) of partial sums $s_n := \sum_{k=0}^{n} c_k$ is convergent (in $\mathbb{C}$).

The series $\sum_{k=0}^{\infty} c_k$ is *absolutely convergent*, if the (real) series $\sum_{k=0}^{\infty} |c_k|$ converges (in $\mathbb{R}$).

PROPOSITION: (i) Basic comparison test: Let (a_n) be a sequence with $a_n \geq 0$ (thus, real!) and such that $\sum a_n$ is convergent. If (c_n) is a complex sequence with the property

$$\exists n_0 \in \mathbb{N} \; \forall n \geq n_0 : \quad |c_n| \leq a_n,$$

then the series $\sum_{n=0}^{\infty} c_n$ is absolutely convergent.

(ii) The root test and the ratio test both are valid for complex sequences literally as stated in Section 4. In particular, if a complex sequence (c_n) with $c_n \neq 0$ (for almost all n) satisfies

$$\exists \theta \in [0, 1[: \quad \left| \frac{c_{n+1}}{c_n} \right| \leq \theta,$$

then the series $\sum_{n=0}^{\infty} c_n$ is absolutely convergent.

(iii) The Proposition concerning the Cauchy product for absolutely convergent series holds literally as stated in Section 4.

Proof. Can be literally copied from those of the corresponding statements about real series. □

(c) Continuity of functions of a complex variable:

DEFINITION: Let $D \subseteq \mathbb{C}$, $z_0 \in \mathbb{C}$. A function $f\colon D \to \mathbb{C}$ is *continuous* at z_0, if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall z \in D: \quad |z - z_0| < \delta \Longrightarrow |f(z) - f(z_0)| < \varepsilon,$$

or equivalently

$$\forall \varepsilon > 0\ \exists \delta > 0: \quad f(U_\delta(z_0) \cap D) \subseteq U_\varepsilon(f(z_0)).$$

f is said to be continuous on D, if f is continuous at all points in D.

REMARK: As in the case of functions on $\mathbb{R}$, continuity can be tested by sequences (the proof is also a literal translation of that in the real case): $f\colon D \to \mathbb{C}$ is continuous at $w \in D$ if and only if or all sequences (z_n) with $z_n \in D$ and $z_n \to w$ $(n \to \infty)$ we have that $\lim f(z_n) = f(w)$. We also express the latter fact by $\lim\limits_{z \to w} f(z) = f(w)$.

6.6. The complex exponential function:

THEOREM: (i) For all $z \in \mathbb{C}$ the series $\sum\limits_{k=0}^{\infty} \frac{z^k}{k!}$ is absolutely convergent. We thus define the complex exponential function $\exp\colon \mathbb{C} \to \mathbb{C}$ by

$$\exp(z) = e^z := \sum_{k=0}^{\infty} \frac{z^k}{k!} \qquad (z \in \mathbb{C}).$$

When restricted to $\mathbb{R}$ it coincides with the exponential function defined in Section 4. We will continue to use the same notation for both functions.

(ii) For all $N \in \mathbb{N}$

$$\exp(z) = \sum_{k=0}^{N} \frac{z^k}{k!} + R_{N+1}(z),$$

where

$$|R_{N+1}(z)| \le 2\frac{|z|^{N+1}}{(N+1)!} \qquad (z \in \mathbb{C}, |z| \le 1 + \frac{N}{2}).$$

(iii) We have the functional equation

$$\forall z_1, z_2 \in \mathbb{C}: \quad \exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2).$$

(iv) For all $z \in \mathbb{C}$: $\exp(\overline{z}) = \overline{\exp(z)}$.

(v) For all $z \in \mathbb{C}$: $\exp(z) \neq 0$.

(vi) $\displaystyle\lim_{z\neq 0, z\to 0} \frac{e^z - 1}{z} = 1$.

(vii) $\exp\colon \mathbb{C} \to \mathbb{C}$ is continuous (at all points of $\mathbb{C}$).

Proof. (i): If $z = 0$ the assertion is trivial. If $z \neq 0$ we apply the ratio test with $c_n = z^k/k!$. For all $n \geq 2|z|$ we have

$$\left|\frac{c_{n+1}}{c_n}\right| = \left|\frac{z^{n+1} n!}{z^n (n+1)!}\right| = \frac{|z|}{n+1} \leq \frac{1}{2} < 1,$$

which proves absolute convergence.

If we temporarily use the notation $\exp_{\mathbb{R}}$ for the (real) exponential function defined in Section 4, then for $x \in \mathbb{R}$ we have $\exp(x + i0) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = \exp_{\mathbb{R}}(x)$.

(ii), (iii): Literally as in the corresponding proofs in Section 4.

(iv): Let $s_n(z) := \sum_{k=0}^{n} z^k/k!$ and use 6.5(a): $\exp(\overline{z}) = \lim s_n(\overline{z}) = \overline{\lim s_n(z)} = \overline{\exp(z)}$.

(v): The functional equation gives $\exp(z)\exp(-z) = \exp(z - z) = \exp(0) = 1$, hence $\exp(z) \neq 0$.

(vi): By (ii) we have $|e^z - 1 - z| \leq 2\frac{|z|^2}{2} = |z|^2$ for all $|z| \leq 3/2$, hence

$$\left|\frac{e^z - 1}{z} - 1\right| \leq |z| \to 0 \quad (z \to 0).$$

(vii): By (vi) we have $e^z - 1 = o(z)$ $(z \to 0)$. Therefore $\lim_{z\to 0} \exp(z) = 1 = \exp(0)$, which shows continuity of exp at 0.

Let $w \in \mathbb{C}$ arbitrary and assume that (z_n) is a sequence in $\mathbb{C}$ such that $z_n \to w$ $(n \to \infty)$. Then $z_n - w \to 0$, thus

$$1 = \exp(0) = \lim_{n\to\infty} \exp(z_n - w) = \lim_{n\to\infty} \exp(z_n)\exp(-w),$$

which implies that $\lim \exp(z_n) = \exp(w)$, hence the continuity of exp at w. $\square$

6.7. Trigonometric functions ⟨*trigonometrische Funktionen oder Winkelfunktionen*⟩[4]:

DEFINITION: We define the *cosine (function)* ⟨*Cosinus (Funktion)*⟩ by

$$\cos\colon \mathbb{R} \to \mathbb{R}, \quad \cos(x) := \operatorname{Re}(\exp(ix)) = \operatorname{Re}(e^{ix}),$$

and the *sine (function)* ⟨*Sinus (Funktion)*⟩ by

$$\sin\colon \mathbb{R} \to \mathbb{R}, \quad \sin(x) := \operatorname{Im}(\exp(ix)) = \operatorname{Im}(e^{ix}).$$

BASIC PROPERTIES: (i) Since $e^{ix} = \operatorname{Re}(e^{ix}) + i\operatorname{Im}(e^{ix})$ we obtain *Euler's formula*

$$\forall x \in \mathbb{R}: \qquad e^{ix} = \cos(x) + i\sin(x). \tag{6.3}$$

Furthermore, cos and sin are continuous $\mathbb{R} \to \mathbb{R}$, since $\exp(ix_n) \to \exp(ia)$ if and only if $\operatorname{Re}(\exp(ix_n)) \to \operatorname{Re}(\exp(ia))$ and $\operatorname{Im}(\exp(ix_n)) \to \operatorname{Im}(\exp(ia))$.

(ii) Geometric interpretation: Since any real t gives $|e^{it}|^2 = e^{it} \cdot \overline{(e^{it})} = e^{it}e^{-it} = e^0 = 1$, we obtain

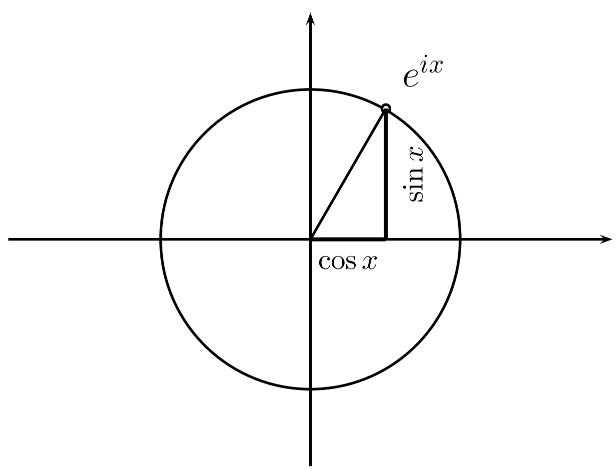
$$|e^{it}| = 1 \qquad \forall t \in \mathbb{R}.$$

Therefore every number of the form e^{it} lies on the unit circle

$$S^1 := \{z \in \mathbb{C} : |z| = 1\} \cong \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$$

and $(\cos(t), \sin(t))$ represents the (Cartesian) coordinates in the plane. In particular, we have the relation

$$\cos^2(x) + \sin^2(x) = 1 \qquad \forall x \in \mathbb{R}. \tag{6.4}$$



[4]These so-called trigonometric functions have a very long history: They were first used by the Babylonians in around 1900 BC and later in the Hellenistic world, in medieval India, in the Islamic Persia and in the medieval Europe. The terms sine and cosine (from the Latin sinus, i.e. "arch ") were introduced by the German mathematician Georg von Peuerbach.

Remark: Note that we avoided any reference to notions like arc length ⟨*Bogenlänge*⟩ or angle ⟨*Winkel*⟩ in defining the trigonometric functions for reasons of a deductive presentation. Arc length will be introduced rigorously, and in more generality, later during the course (based on the notion of integrals along curves), but it certainly is useful to have the intuitive meaning at hand already as suggested by the above geometric interpretation.

(iii) Recall that for any complex number w the real and imaginary part can be obtained from $\mathrm{Re}(w) = (w + \overline{w})/2$ and $\mathrm{Im}(w) = (w - \overline{w})/2i$. Therefore we have

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i},$$

which in turn implies

$$\cos(-x) = \cos(x) \quad \text{and} \quad \sin(-x) = -\sin(x),$$

telling that cos is an even ⟨*gerade*⟩ function (the graph is symmetric with respect to the vertical axis) and sin is an odd ⟨*ungerade*⟩ function (the graph is reflected by lines through the origin $(0, 0)$).

(iv) The fundamental relations for the addition of arguments ("angles") ⟨*Additionstheoreme*⟩ are the following: For all $x, y \in \mathbb{R}$

$$\begin{aligned} \cos(x + y) &= \cos(x)\cos(y) - \sin(x)\sin(y) \\ \sin(x + y) &= \cos(x)\sin(y) + \sin(x)\cos(y) \end{aligned}$$

and

$$\begin{aligned} \cos(x) - \cos(y) &= -2\sin\frac{x+y}{2}\sin\frac{x-y}{2} \\ \sin(x) - \sin(y) &= 2\cos\frac{x+y}{2}\sin\frac{x-y}{2}. \end{aligned}$$

Proof. The first two equations are obtained by taking real and imaginary parts in the relation

$$e^{i(x+y)} = e^{ix} \cdot e^{iy}.$$

The third (resp. fourth) equation follows from the first (resp. second) equation upon setting $u = (x + y)/2$ and $v = (x - y)/2$ ($\Leftrightarrow x = u + v,\ y = u - v$):

$$\begin{aligned} \cos(x) - \cos(y) &= \cos(u + v) - \cos(u - v) \\ &= \cos(u)\cos(v) - \sin(u)\sin(v) - \big(\cos(u)\cos(v) + \sin(u)\sin(v)\big) \\ &= -2\sin(u)\sin(v) = -2\sin\frac{x+y}{2}\sin\frac{x-y}{2}, \end{aligned}$$

and similarly for the last equation. □

(v) The natural integer powers of i show a simple repetitive pattern: Since $i^2 = -1$, $i^3 = i^2 i = -i$, $i^4 = i^3 i = -i^2 = 1$, we have for $n \in \mathbb{N}$ that

$$i^n = \begin{cases} 1 & \text{if } n = 4m \text{ for some } m \in \mathbb{N} \quad (\Leftrightarrow n \equiv 0 \mod 4) \\ i & \text{if } n = 4m+1 \text{ for some } m \in \mathbb{N} \quad (\Leftrightarrow n \equiv 1 \mod 4) \\ -1 & \text{if } n = 4m+2 \text{ for some } m \in \mathbb{N} \quad (\Leftrightarrow n \equiv 2 \mod 4) \\ -i & \text{if } n = 4m+3 \text{ for some } m \in \mathbb{N} \quad (\Leftrightarrow n \equiv 3 \mod 4). \end{cases}$$

Therefore we obtain for all $x \in \mathbb{R}$

$$\begin{aligned} \cos(x) + i\sin(x) = e^{ix} = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = \sum_{n=0}^{\infty} \frac{i^n x^n}{n!} \\ = \underbrace{\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}}_{\mathrm{Re}(e^{ix})} + i \cdot \underbrace{\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}}_{\mathrm{Im}(e^{ix})}, \end{aligned}$$

which proves the following series expansions for cosine and sine:

$$\cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}, \qquad \sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

(vi) Suppose we are to use the above series expansions to approximate cosine and sine for small x by simply dropping all terms that contain x to quadratic or higher order. If justified, this would give the following simple heuristic relations when $|x|$ is small:

$$\cos(x) \approx 1 \quad \text{and} \quad \sin(x) \approx x.$$

As a matter of fact, we have the limit equations

$$\lim_{x \neq 0, x \to 0} \frac{\cos(x) - 1}{x} = 0, \qquad \lim_{x \neq 0, x \to 0} \frac{\sin(x)}{x} = 1.$$

Proof. By Theorem 6.6(vi) we have that

$$1 + i \cdot 0 = \lim_{x \to 0} \frac{e^{ix} - 1}{ix} = \lim_{x \to 0} \mathrm{Re}(\frac{e^{ix} - 1}{ix}) + i \cdot \lim_{x \to 0} \mathrm{Im}(\frac{e^{ix} - 1}{ix}).$$

Therefore we have for $x \in \mathbb{R}$ as $x \to 0$

$$\begin{aligned} \frac{\cos(x) - 1}{x} &= -\mathrm{Im}(\frac{\cos(x) - 1 + i\sin(x)}{ix}) = -\mathrm{Im}(\frac{e^{ix} - 1}{ix}) \to 0 \\ \frac{\sin(x)}{x} &= \mathrm{Re}(\frac{\cos(x) - 1 + i\sin(x)}{ix}) = \mathrm{Re}(\frac{e^{ix} - 1}{ix}) \to 1. \end{aligned}$$

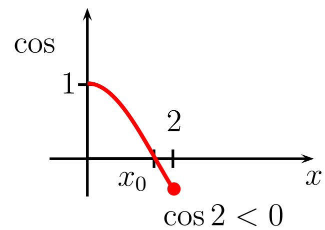
□

6.8. Definition of π: We will show that cos is strictly decreasing on the interval $[0,2]$ and possesses a unique zero x_0 in that interval. We will define π as the value of $2x_0$. We postpone the precise identification of π with the (length of the) circumference of the unit circle until integration theory allows us to provide a simple calculation.[5]

LEMMA: (i) $\cos(0) = 1$ and $\cos(2) \leq -1/3$.

(ii) If $0 < x \leq 2$ then $\sin(x) > 0$.

(iii) cos is strictly decreasing on $[0, 2]$.



Proof. (i): We clearly have $\cos(0) = \mathrm{Re}(e^{i0}) = 1$.

The series expansion for the cosine function gives the alternating sum

$$\cos(2) = 1 - \frac{2^2}{2!} + \sum_{k=2}^{\infty}(-1)^k \frac{2^{2k}}{(2k)!} = -1 + r,$$

where r represents the error when approximating $\cos(2)$ by the partial sum $s_1 = -1$. Thus the error estimate (4.2) from the Leibniz criterion tells that $|r|$ is bounded by the absolute value of the first neglected term. Therefore we have

$$\cos(2) \leq -1 + |r| \leq -1 + \frac{2^4}{4!} = -1 + \frac{16}{24} = -1 + \frac{2}{3} = -\frac{1}{3}.$$

(ii): Let $0 < x \leq 2$. We have the alternating sum for the sine function

$$\sin(x) = x + \sum_{k=1}^{\infty}(-1)^k \frac{x^{2k+1}}{(2k+1)!} = x + r(x),$$

where $r(x)$ now denotes the error when approximating $\sin(x)$ by the partial sum $s_1(x) = x$. We apply again the estimate (4.2), which now reads

$$|r(x)| \leq \frac{x^3}{3!} = x \cdot \frac{x^2}{6} \leq x \cdot \frac{4}{6} = \frac{2x}{3},$$

and therefore

$$\sin(x) \geq x - |r(x)| \geq x - \frac{2x}{3} = \frac{x}{3} > 0.$$

(iii): Let $0 < x_1 < x_2 \leq 2$, then $0 < (x_1 + x_2)/2 \leq 2$ as well as $0 < (x_2 - x_1)/2 \leq 2$. By 6.7(iv) and property (ii) proved above

[5] This constant was first named "π" by the Welsh scientist William Jones in 1706 because it is the first letter of the Greek words περιφερεία ("periphery") and περίμετρος ("circumference"). This notation was later adopted by Euler.

we obtain

$$\cos(x_2) - \cos(x_1) = -2 \cdot \underbrace{\sin\frac{x_2 + x_1}{2}}_{>0} \cdot \underbrace{\sin\frac{x_2 - x_1}{2}}_{>0} < 0,$$

hence $\cos(x_2) < \cos(x_1)$. □

PROPOSITION: There exists a unique $x_0 \in [0, 2]$ such that $\cos(x_0) = 0$.

Proof. By the above lemma, cos is strictly decreasing on $[0, 2]$, hence $\cos|_{[0,2]}$ is injective. Furthermore, the same lemma gives that $\cos(0) > 0$ and $\cos(2) < 0$. Since cos is continuous, the intermediate value theorem implies the existence of a zero $x_0 \in [0, 2]$. This zero must be unique, since cos is injective on that interval. □

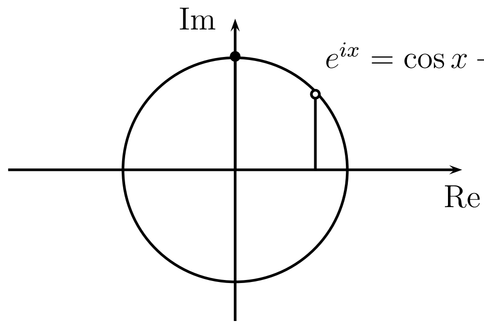
DEFINITION: Let x_0 denote the unique zero of cos in the interval $[0, 2]$ (according to the above proposition). Then the real number π is defined by $\pi := 2x_0$.

The properties of cos and sin established above can now be reformulated in more familiar terms: For example, we obtain that

$$\cos(x) > 0 \quad \text{for } 0 \leq x < \frac{\pi}{2}, \quad \cos(\frac{\pi}{2}) = 0, \quad \cos(x) < 0 \quad \text{for } \frac{\pi}{2} < x \leq 2.$$

Since $\sin^2(\frac{\pi}{2}) = 1 - \cos^2(\frac{\pi}{2}) = 1$ and $\sin(\frac{\pi}{2}) > 0$ (by the above lemma), we have

$$\sin(\frac{\pi}{2}) = 1 \quad \text{and} \quad e^{i\frac{\pi}{2}} = \cos(\frac{\pi}{2}) + i\sin(\frac{\pi}{2}) = i.$$



Taking integer powers for all $k \in \mathbb{Z}$ we obtain $e^{ik\frac{\pi}{2}} = \left(e^{i\frac{\pi}{2}}\right)^k = i^k$. In particular,

$$e^{i\,0} = 1 = \cos(0) + i\sin(0), \quad e^{i\frac{\pi}{2}} = i = \cos(\frac{\pi}{2}) + i\sin(\frac{\pi}{2}), \quad e^{i\pi} = -1 = \cos(\pi) + i\sin(\pi),$$
$$e^{i\frac{3\pi}{2}} = -i = \cos(\frac{3\pi}{2}) + i\sin(\frac{3\pi}{2}), \quad e^{i2\pi} = 1 = \cos(2\pi) + i\sin(2\pi),$$

which is summarized in the table

x	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
$\sin(x)$	0	1	0	-1	0
$\cos(x)$	1	0	-1	0	1

6.9. Further properties of the trigonometric functions: For all $x \in \mathbb{R}$ we have the following properties:

(a) cos and sin are *periodic* ⟨periodisch⟩ with *period* ⟨Periode⟩ of 2π, i.e.,

$$\cos(x + 2\pi) = \cos(x), \qquad \sin(x + 2\pi) = \sin(x).$$

This follows from 6.7(iv) and the fact that $\cos(2\pi) = 1$, $\sin(2\pi) = 0$:

$$\cos(x + 2\pi) = \cos(x)\cos(2\pi) - \sin(x)\sin(2\pi) = \cos(x).$$

(b) Since $\cos(x + \pi) = \cos(x)\cos(\pi) - \sin(x)\sin(\pi) = -\cos(x)$, and a similar calculation for sin, we have

$$\cos(x + \pi) = -\cos(x), \qquad \sin(x + \pi) = -\sin(x).$$

(c) By $\sin(\frac{\pi}{2} - x) = \sin(\frac{\pi}{2})\cos(-x) + \cos(\frac{\pi}{2})\sin(-x) = \cos(x)$, and a similar calculation for cos, we obtain

$$\sin(\frac{\pi}{2} - x) = \cos(x), \qquad \cos(\frac{\pi}{2} - x) = \sin(x).$$

(d) $\sin(x) = 0 \iff x \in \pi\mathbb{Z} := \{k\pi : k \in \mathbb{Z}\}$

Proof. By 2π-periodicity it suffices to show the assertion for $x \in [0, 2\pi[$.

Let $0 < x < \pi$ arbitrary, then $\frac{\pi}{2} - x \in]-\frac{\pi}{2}, \frac{\pi}{2}[$ and therefore $\sin(x) = \cos(\frac{\pi}{2} - x) > 0$.

Furthermore, note that $]\pi, 2\pi[= \{r + \pi : 0 < r < \pi\}$ and $\sin(x + \pi) = -\sin(x) < 0$.

Thus, 0 and π are the only zeros of sin in the interval $[0, 2\pi[$, which proves the assertion. □

(e) $\cos(x) = 0 \iff x \in \{\frac{\pi}{2}\} + \pi\mathbb{Z} := \{\frac{\pi}{2} + k\pi : k \in \mathbb{Z}\}$

Proof. Use $\cos(x) = -\sin(x - \frac{\pi}{2})$ and apply (d). □

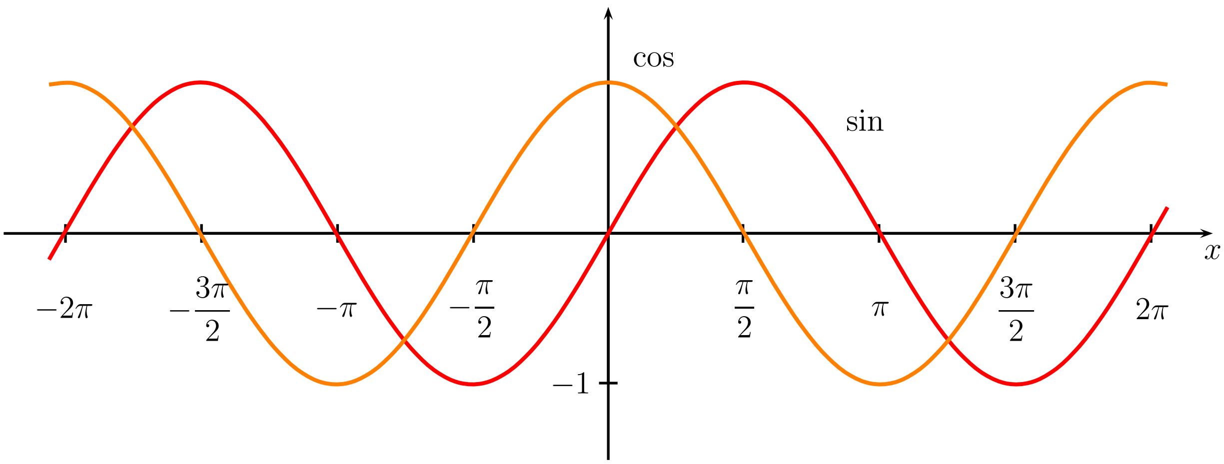
(f) $e^{ix} = 1 \iff x \in 2\pi\mathbb{Z} := \{2k\pi : k \in \mathbb{Z}\}$

Proof. We have $e^{ix} - 1 = e^{i\frac{x}{2}} \cdot \left(e^{i\frac{x}{2}} - e^{-i\frac{x}{2}}\right) = 2i\, e^{i\frac{x}{2}} \sin\frac{x}{2}$ and $2i\, e^{i\frac{x}{2}} \neq 0$. Therefore

$$e^{ix} = 1 \iff \sin(\frac{x}{2}) = 0 \iff \frac{x}{2} \in \pi\mathbb{Z} \iff x \in 2\pi\mathbb{Z}.$$

□

Using the above list of basic properties of cos and sin we can get a good qualitative picture of their graphs. Note in particular the following features: a shift of the graph of cos by $\frac{\pi}{2}$ along the horizontal axes gives the graph of sin; cos is even and strictly decreasing on $[0, \pi]$ (thus increasing on $[-\pi, 0]$), sin is odd and strictly increasing on $[-\frac{\pi}{2}, \frac{\pi}{2}]$; besides the zeros we can read off locations of maximum and minimum values, where each functions changes monotonicity type from increasing to decreasing or vice versa.



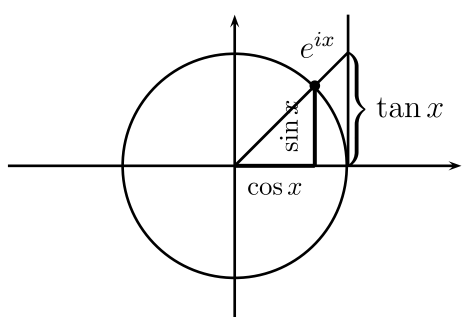
DEFINITION: (i) The *tangent (function)*[6] ⟨*Tangens (-Funktion)*⟩ $\tan\colon \mathbb{R}\setminus(\frac{\pi}{2}+\pi\mathbb{Z})\to\mathbb{R}$ is given by

$$\tan(x) := \frac{\sin(x)}{\cos(x)}.$$

(ii) The *cotangent (function)* ⟨*Cotangens (-Funktion)*⟩ $\cot\colon \mathbb{R}\setminus\pi\mathbb{Z}\to\mathbb{R}$ is given by

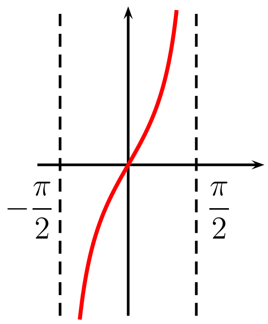
$$\cot(x) := \frac{\cos(x)}{\sin(x)}.$$

A geometric interpretation of $\tan(x)$, when $-\frac{\pi}{2} < x < \frac{\pi}{2}$, is easy by comparing the right-angled triangles in the following illustration:



[6]The term tangent was first used by the Danish mathematician Thomas Fincke in 1583.

Here is the part of the graph of tan above the interval $]-\frac{\pi}{2},\frac{\pi}{2}[$, whose basic qualitative features can be derived from the properties of cos and sin:

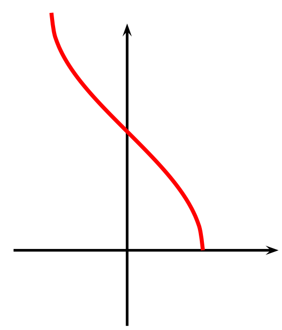


Note that $\tan(x+\pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{-\sin(x)}{-\cos(x)} = \tan(x)$, so that the complete graph of tan can be obtained from shifts of the basic part on $]-\frac{\pi}{2},\frac{\pi}{2}[$ by integer multiples of π.

6.10. Inverse trigonometric functions ⟨*Arcusfunktionen*⟩:

Arc cosine: We assert that cos is strictly decreasing on $[0,\pi]$ and $\cos([0,\pi]) = [-1,1]$.

Indeed, that cos is strictly decreasing on $[0,\frac{\pi}{2}]$ follows from the Lemma in 6.8; since $\cos(\pi - x) = -\cos(x)$ the same follows for the interval $[\frac{\pi}{2},\pi]$; by continuity and injectivity, $\cos([0,\pi]) = [\cos(\pi),\cos(0)] = [-1,1]$.



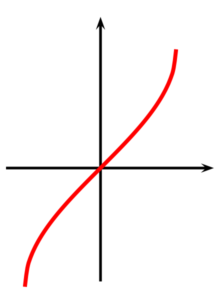
Thus cos is continuous, strictly decreasing, and bijective as a map $[0,\pi]\to[-1,1]$, hence possesses a strictly decreasing continuous inverse function

$$\arccos\colon [-1,1]\to[0,\pi],$$

called the *arc cosine (function)* ⟨*Arcus Cosinus*⟩.

We have for all $x\in[0,\pi]$ that $\arccos(\cos(x)) = x$ and $\cos(\arccos(y)) = y$ for all $y\in[-1,1]$.

Of course we could have constructed similar inverses on any interval of strict monotonicity for cos. Unless stated otherwise we will usually refer to the one constructed above as arccos.



Arc sine: Using $\sin(x) = \cos(\frac{\pi}{2}-x)$ and (i) we deduce the following: sin is strictly increasing on $[-\frac{\pi}{2},\frac{\pi}{2}]$ and $\sin([-\frac{\pi}{2},\frac{\pi}{2}]) = [-1,1]$. The corresponding inverse function

$$\arcsin\colon [-1,1]\to[-\frac{\pi}{2},\frac{\pi}{2}],$$

called the *arc sine (function)* ⟨*Arcus Sinus*⟩, is continuous and strictly increasing.

Arc tangent: We claim that tan is strictly increasing on $]-\frac{\pi}{2}, \frac{\pi}{2}[$ and $\tan(]-\frac{\pi}{2}, \frac{\pi}{2}[) = \mathbb{R}$.

Proof. Since

$$\tan(-x) = \frac{\sin(-x)}{\cos(-x)} = \frac{-\sin(x)}{\cos(x)} = -\tan(x),$$

it suffices to consider the subinterval $[0, \frac{\pi}{2}[$. If $0 \leq x < x' < \frac{\pi}{2}$ then $\sin(x) < \sin(x')$ and $\cos(x) > \cos(x') > 0$, hence

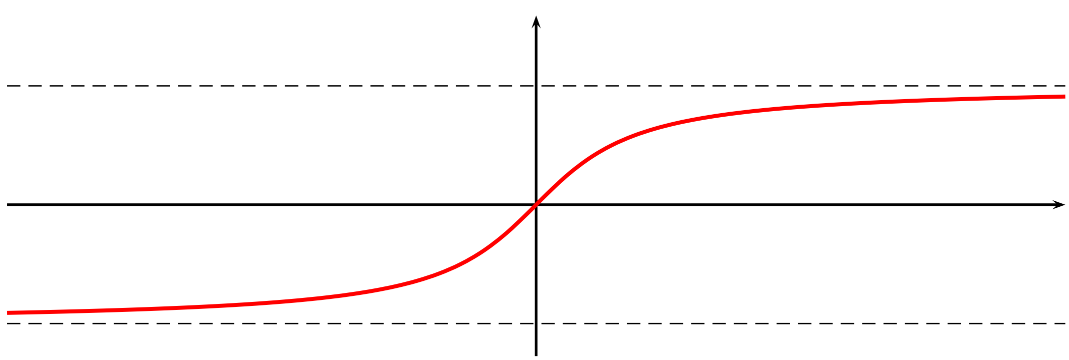
$$\tan(x) = \frac{\sin(x)}{\cos(x)} < \frac{\sin(x')}{\cos(x')} = \tan(x').$$

Note that $\frac{\cos(x)}{\sin(x)} > 0$ for all $x \in]0, \frac{\pi}{2}[$ and that

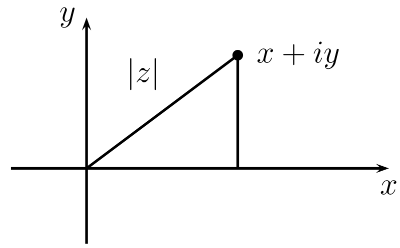
$$\lim_{x \nearrow \frac{\pi}{2}} \frac{\cos(x)}{\sin(x)} = \frac{\cos(\frac{\pi}{2})}{\sin(\frac{\pi}{2})} = 0.$$

Therefore we obtain that $\tan(x) \to \infty$ as $x \nearrow \frac{\pi}{2}$ and by the intermediate value theorem (tan is continuous!) that $\tan([0, \frac{\pi}{2}[) = [0, \infty[$. By symmetry of tan we obtain that $\tan(]-\frac{\pi}{2}, \frac{\pi}{2}[) =]-\infty, \infty[$. □

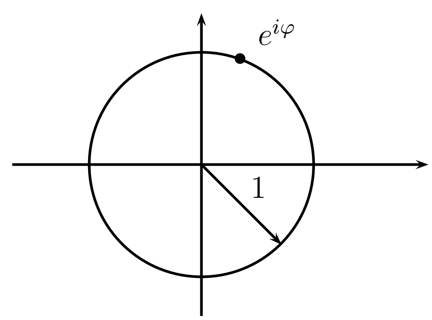
We conclude that the restriction of tan to $]-\frac{\pi}{2}, \frac{\pi}{2}[$ has a continuous and strictly increasing inverse function $\arctan\colon \mathbb{R} \to]-\frac{\pi}{2}, \frac{\pi}{2}[$, called *arc tangent (function)* ⟨*Arcus Tangens*⟩.



6.11. Polar coordinates[7] for complex numbers: If $z = x + iy \in \mathbb{C}$ we may interpret the absolute value $|z| = \sqrt{x^2 + y^2}$ as the distance of z to the origin in the plane:



How do we obtain information on the direction towards z as seen from the origin with respect to the positive real axis (the x-axis)?



Recall that for any $\varphi \in \mathbb{R}$ we have $|e^{i\varphi}| = 1$ and $e^{i\varphi} = \cos(\varphi) + i\,\sin(\varphi)$.

Furthermore, the points where the unit circle S^1 intersects the Cartesian axes are given by $e^{i0} = 1$, $e^{i\frac{\pi}{2}} = i$, $e^{i\pi} = -1$, $e^{i\frac{3\pi}{2}} = -i$ and we have 2π-periodicity

$$e^{i(\varphi+2\pi)} = e^{i\varphi}.$$

Let $z \neq 0$ and set $w := \frac{z}{|z|}$. Then w lies on the unit circle and can be written in the form

$$w = \xi + i\,\eta, \quad \text{where } \xi, \eta \in \mathbb{R} \text{ such that} \quad 1 = |w|^2 = \xi^2 + \eta^2.$$

Therefore $\xi \in [-1, 1]$ and $\alpha := \arccos(\xi) \in [0, \pi]$ and we have

$$\sin^2(\alpha) = 1 - \cos^2(\alpha) = 1 - \xi^2 = \eta^2,$$

hence $\sin(\alpha) = \eta$ or $\sin(\alpha) = -\eta$. If we define

$$\varphi := \begin{cases} \alpha & \text{if } \sin(\alpha) = \eta \\ -\alpha & \text{if } \sin(\alpha) = -\eta, \end{cases}$$

then we obtain $w = \cos(\varphi) + i\,\sin(\varphi) = e^{i\varphi}$, which in turn yields the *polar representation* ⟨*Polardarstellung*⟩ of z in the form

$$z = |z| \cdot e^{i\varphi}.$$

In this representation φ, the so-called the *argument* ⟨*Argument*⟩ of z, $\varphi = \arg(z)$, is unique up to an addition of integer multiples of 2π.

[7] The history of polar coordinates is about as long as the one of trigonometry. The term polar coordinates was introduced by 18th-century Italian mathematicians.

BIBLIOGRAPHY

[AE02] H. Amann and J. Escher. *Analysis I*. Birkhäuser, Basel, 2002. 2. Auflage.

[BF00] M. Barner and F. Flohr. *Analysis I*. Walter de Gruyter, Berlin, 2000. 5. Auflage.

[Heu88] H. Heuser. *Analysis 1*. BG Teubner, Stuttgart, 1988. 5. Auflage.

[SS07] H. Schichl and R. Steinbauer. *Einführung in das mathematische Arbeiten*. Universität Wien, 2007. Vorlesungsskriptum.